

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESPRIT SANS FRONTIÈRES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN PHILOSOPHIE

PAR

LOUIS CHARTRAND

NOVEMBRE 2014

REMERCIEMENTS

Une des conséquences de l'esprit étendu est que la responsabilité du travail cognitif, incluant le travail intellectuel, peut s'étendre au-delà de celle ou celui qui en prend traditionnellement l'autorité. C'est particulièrement vrai dans le cas de ce travail qui, en toute justice, est véritablement un travail de groupe.

Il ne me serait pas possible de mentionner toutes celles et ceux qui ont contribué d'une façon ou d'une autre à ce travail. Cependant, quelques remerciements particuliers s'imposent. En premier lieu, je veux remercier les gens du Laboratoire d'analyse cognitive de l'information (LANCI), à l'UQAM, qui m'ont donné un milieu de recherche incomparable. Mes collègues et professeurs ont contribué à ce qu'il y a de mieux dans ce mémoire, en idées, en apprentissages et à travers de nombreuses discussions. Je tiens spécialement à remercier Sarah Arnaud, Anne Brel Cloutier, Jean-François Chartier, Jean-Guy Meunier, Pierre Poirier, Serge Robert et Maxime Sainte-Marie. Je tiens également à remercier ceux et celles qui ont généreusement offert de relire mon texte et m'ont proposé des corrections éclairantes et inspirantes : Arielle Acosta Perelman, Jean-Guy Meunier, Pierre Poirier, Luce Vermette et Natalie Voarino.

Un tel projet ne serait pas possible sans le soutien de mes proches, qui ont vécu les moments les plus difficiles avec moi et m'ont épaulé. De nombreuses personnes mériteraient mention, mais je remercierai particulièrement mes parents, Luce Vermette et René Chartrand, mon frère Alexandre, mes amis Yohan Cormier, et Geneviève Dick, et mon amoureuse Audrey Rousseau.

TABLE DES MATIÈRES

Remerciements	iii
Résumé	ix
CHAPITRE I	
INTRODUCTION	1
1.1 Explorer l'esprit	3
1.1.1 Représentation : le berceau de l'esprit étendu	4
1.1.2 L'absence de théorie du cognitif	7
1.2 La question	19
1.2.1 Caractérisation et extension	19
1.2.2 Problématique	22
1.3 La démarche	26
1.4 Analyse de texte	28
1.5 Théorie de la caractérisation	30
1.6 Usages du concept d'esprit – qui en a besoin ?	35
1.7 Conclusion	37
CHAPITRE II	
REPOUSSER LA QUESTION	39
2.1 Critères de sélection	40
2.2 Le principe de parité	44
2.2.1 Quatre versions du principe de parité	45

2.2.2	L'intuition internaliste et les Martiens	48
2.2.3	Fonctionnalisme et thermostat humain	52
2.3	Manipulabilité mutuelle	55
2.4	L'accord avec l'évidence empirique	59
2.4.1	Traitement de l'évidence empirique chez Clark	61
2.4.2	Traitement de l'évidence empirique chez Rupert	67
2.4.3	Empiriquement indistinguable	69
2.5	Le contenu non-dérivé	74
2.6	Conclusion	77
CHAPITRE III		
INDIVIDUALISME		79
3.1	Méthodologie	79
3.2	Intuition individualiste et frontières	81
3.2.1	La critique du schéma entrée-sortie	82
3.2.2	L'insaisissable légèreté du lien entre niveaux	85
3.2.3	Frontières	86
3.2.4	Détecter l'intuition individualiste	88
3.3	Rupert	93
3.4	Clark	97
3.5	Haugeland	101
3.6	Rowlands	102
3.7	Conclusion	106
CHAPITRE IV		
CONCLUSION		109
4.1	Bilan de l'argumentation	109
4.1.1	Isomorphie	110
4.1.2	Caractère individuel	113
4.1.3	Bilan des hypothèses	120
4.2	Horizons	122

4.2.1	Sciences humaines et intuition individualiste	122
4.2.2	Contribution aux sciences cognitives	125
	Références	140

RÉSUMÉ

La plupart des auteur·es ayant abordé le problème de l'extension du cognitif, tel qu'il a émergé des débats autour de la thèse de l'esprit étendu, ont supposé que cette extension devait prendre la forme d'un espace régulier, qui peut être ceint par des frontières. Cependant, la littérature en question ne traite pas explicitement de cette supposition, de sorte que, malgré son influence, il n'y a pas d'évaluation de sa véracité ou de sa légitimité. Dans ce mémoire, cette hypothèse est remise en question, et nous prenons sa contrepartie pour hypothèse de recherche, à savoir que, malgré l'assurance des auteur·es qui l'épousent, il est tout à fait légitime d'explorer des propositions de caractérisation du cognitif qui aboutissent à d'autres formes d'extension. Pour démontrer notre hypothèse (H1), nous posons deux autres hypothèses : (H2) les auteur·es qui épousent la supposition selon laquelle l'esprit doit s'étendre sur un espace ceint par des frontières le font sur la base de l'intuition individualiste, et (H3) l'intuition individualiste, malgré certaines apparences, ne permet pas de justifier cette supposition. Nous démontrons H2 à l'aide d'une analyse de texte faite sur le corpus des auteur·es ayant abordé la question de l'extension du cognitif dans la foulée des débats sur l'esprit étendu, et nous montrons que la littérature en sciences cognitives et en philosophie de l'esprit nous donne des raisons de croire que H3 est vraie. Ensemble, H2 et H3 justifient H1.

MOTS-CLÉS : sciences cognitives, philosophie de l'esprit, esprit étendu, analyse de texte

CHAPITRE I

INTRODUCTION

L'article que l'on considère désormais comme l'inspiration de toute la littérature sur l'*extended mind* (Clark et Chalmers, 1998) a été refusé trois fois et a attendu trois ans avant d'être finalement publié dans la revue *Analysis* (Baggini, 2008). Mais depuis, selon les mots de son auteur, David Chalmers, « it has taken a life of its own ». Ce n'est pas tant un commentaire quant à la direction qu'a pris le débat – qui, en majeure partie, est resté proche des préoccupations de l'article séminal – que sur l'ampleur que la question a prise dans la philosophie de l'esprit et des sciences cognitives.

Cette ampleur tient à deux choses. D'une part, le concept de l'esprit étendu remet en question une (pré)conception assez bien acceptée par les cognitivistes et autres philosophes de l'esprit, à savoir que le cerveau est le lieu où s'incarne ou se réalise l'esprit. Pour s'en convaincre, il suffit de constater que les débats qui ont eu lieu dans les années 1950 et 1960 autour de la *mind-brain identity thesis* n'avaient pas pour enjeu l'espace où s'incarnerait l'esprit, mais plutôt la possibilité logique d'identifier l'esprit ou les états mentaux avec un principe matériel. Pour ces théoriciens, si cette identité était possible, elle devait se faire avec le cerveau – ça semblait aller de soi (Smart, 2011). Aussi, de nombreux développements en philosophie, en psychologie et en éthique dépendent-ils de cette « évidence ». Sa remise en question pourrait donc affecter bien des disciplines.

D'autre part, si l'argument de « The Extended Mind » a su ébranler bien des certitudes, son objectif n'est pas de formuler des critères qui nous permettent de déterminer l'extension de l'esprit. La littérature qui a suivi s'est avérée abondante, et elle a proposé plusieurs critères, mais on ne peut dire qu'aucun n'ait réussi à s'imposer. Alors que le débat se prolonge et rejoint de nouveaux cercles, des avancées sont faites, mais elles ne semblent que confirmer qu'on n'a pas encore les moyens de départager les différentes positions sur la question (Barker, 2010; Sprevak, 2010). Bref, *the jury is still out*.

Quinze ans après l'acte de naissance officiel de l'esprit étendu, on peut commencer à se demander si le débat ne risque pas de s'étioler. Aussi est-il pertinent de remettre en question notre façon d'aborder le sujet, en commençant par les questions que l'on se pose. Dans les faits, le débat entre partisans et détracteurs de l'esprit étendu porte principalement par l'extension spatiale qu'on lui accorde. Cependant, à force de travailler avec des critères trop faibles pour gagner de consensus, les débats en sont venus à ressembler un peu à un marchandage sur la démarcation entre ce qui est interne et externe à l'esprit. Mais est-il même vraiment pertinent de chercher une démarcation ?

Concevoir l'esprit en des termes qui permettent de comprendre ses mécanismes¹ soulève des difficultés particulières. D'une part, à cause de l'héritage de la tradition dualiste, notre conception de l'esprit repose sur une métaphysique qui a tendance à compliquer le problème. D'autre part, le système cognitif est un système complexe où les interactions causales sont si nombreuses qu'elles sont difficiles à suivre. Aussi est-il normal de se doter d'outils à la hauteur du problème. Il s'agit là de conditions extrêmes pour la conceptualisation, et on ne saurait espérer produire des ontologies et des concepts utiles sans une bonne dose de créativité. Il est donc pertinent de revisiter la façon dont on produit des concepts pour attaquer celui de l'esprit. Or les philosophes ne semblent pas s'être intéressés outre mesure à cette question.

1. Nous entendons « mécanisme » *grosso modo* de la même façon que Machamer *et al.* (2000) – c'est-à-dire comme l'ensemble des causes (physiques) qui expliquent un phénomène. On dira aussi de ces causes qu'elles sont ce qui « réalise » le phénomène

Il y a plusieurs façons de remettre en question l’orthodoxie quant aux façons de concevoir l’esprit, et se questionner sur la pertinence de donner une frontière à l’extension de l’esprit en est une. En fait, si l’hypothèse que nous entendons défendre – à savoir que les mécanismes du cognitif n’ont pas à avoir de frontières définies (plus de détails en section §1.2.2) – s’avère, l’hypothèse selon laquelle les outils traditionnels ne sont pas appropriés pour conceptualiser l’esprit sera elle-même confirmée. Notre entreprise peut donc être une introduction à l’élaboration d’un nouveau régime de pratiques de conceptualisation pour l’esprit et le cognitif. Le choix de faire la critique des frontières, plutôt que d’un autre concept, vient simplement du fait qu’elle nous semble plus accessible et plus sûre, la frontière étant un aspect concret du concept d’esprit. Nous abordons donc la critique du concept d’esprit par ce qui apparaît intuitivement être le chemin le plus facile.

Dans ce premier chapitre, nous élaborons la démarche qui nous guidera pour la suite. Dans un premier temps, nous la situons à l’intérieur de la philosophie de l’esprit et des sciences cognitives, de façon à comprendre ce qui a amené l’extension de l’esprit à occuper la place qu’elle prend dans le paysage intellectuel. Cela nous permettra de mieux comprendre la question qui donne lieu aux débats autour de l’esprit étendu. Ensuite, en sections 1.3 et 1.4, nous formulerons notre hypothèse et nous donnerons un aperçu de notre démarche. Enfin, dans les deux dernières sections, nous aborderons des problèmes particuliers dans cette démarche, qui apparaîtront dans la suite.

1.1 Explorer l’esprit

L’esprit étendu n’est pas une théorie complètement nouvelle – John Dewey (1896, 1916), F.H. Bradley (1895) et G.H. Mead (1938, 1929), notamment, ont épousé des thèses semblables. Cependant, son retour dans les débats philosophiques ne saurait s’expliquer par l’influence de ces philosophes. Il faut donc que les développements récents en philosophie et en sciences cognitives aient créé les conditions pour cette résurgence. Au moins deux facteurs conspiraient à rendre les débats sur

l'extension de l'esprit particulièrement pertinents : l'arrivée de nouveaux modèles de la représentation, et l'absence de caractérisation propre de ce qui constitue un esprit. Nous arguerons ici que ce contexte fait de la question de l'extension de l'esprit un substitut pour la question de la caractérisation du concept d'esprit.

1.1.1 Représentation : le berceau de l'esprit étendu

L'anti-représentationnalisme est une thèse qui a pris une certaine importance au début des années 1990, notamment grâce à la défense qu'en a fait Tim Van Gelder (1990, 1995).

On sait depuis Anscombe (1957) qu'un même système peut recevoir plusieurs descriptions différentes reposant sur des ontologies différentes – par exemple, on conçoit généralement qu'une description psychologique peut se faire en terme d'attitudes propositionnelles (désirs, intentions, croyances) ou en termes plus mécanistes (Davidson, 1963). Van Gelder se base sur cette prémisse, contrastant une description computationnelle en terme d'attitudes propositionnelles à une description en terme de système(s) dynamique(s) – il en donne un exemple avec le gouverneur de Watt. Cependant, argue-t-il, toutes les descriptions ne sont pas égales en regard de l'explication – dans le cas du gouverneur de Watt, la description computationnelle ne présente tout simplement pas l'information pertinente pour comprendre le fonctionnement du système dans toute sa subtilité. La description en terme de système dynamique, en revanche, nous permet de le comprendre beaucoup mieux.

Van Gelder fait le pari qu'il en va de même pour la cognition que pour le gouverneur de Watt. Il illustre son argument sur des modèles de prise de décision : la description computationnelle a du mal à rendre compte de la dimension temporelle de la prise de décision, alors que l'approche dynamiciste y parvient beaucoup mieux. Or cette approche n'emploie pas de représentation. Extrapolant à partir de cet exemple, Van Gelder croit que les descriptions computationnelles ne

conviennent pas à des systèmes complexes, et en conclut que la description (et donc aussi l'explication) de la cognition peut avantageusement se passer de représentations.

Cette critique fait écho, en quelque sorte, à d'autres inquiétudes, notamment celle de Harnad (1990), qui note que le modèle computationnaliste ne permet pas de décrire l'ancrage sensorimoteur nécessaire aux représentations pour qu'elles représentent effectivement ce qu'elles doivent représenter – ce qu'une approche dynamiciste peut très bien décrire. De plus, van Gelder jouit aussi d'une certaine validation dans le domaine empirique, alors que plusieurs travaux semblent confirmer ses hypothèses, notamment ceux de Brooks (1991) et de Hutchins (1991, 1995). Ces facteurs aidant, l'anti-représentationnalisme semble alors avoir le vent dans les voiles, et on commence à considérer le dynamicisme comme une alternative au computationnalisme.

Cependant de nombreux programmes de recherche ayant eu beaucoup de succès comptent sur ces représentations. Faut-il tout réécrire ces succès empiriques en termes dynamicistes pour conserver les acquis de la psychologie cognitive ? Ou peut-on repenser la représentation en des termes qui répondent adéquatement aux critiques des anti-représentationnalistes ?

En réponse aux anti-représentationnalistes, des philosophes choisissent plutôt de modifier considérablement la théorie de la représentation traditionnelle, et proposent de nouveaux modèles (Millikan, 1996; Clark et Toribio, 1994; Clark, 1998; Wheeler, 2005; Rowlands, 2006). Aux représentations passives, discrètes et clairement identifiables du computationnalisme, les nouveaux représentationnalistes substituent de représentations étendues dans l'espace et dans le temps, sans contours précis, et dont la réalisation se fait à travers les mécanismes qui sont responsables de l'action. Les nouvelles représentations – appelons-les *représentations incarnées*² – sont donc intimement dépendantes de leur contexte d'implémentation, de sorte qu'à travers l'action, le lien sensorimoteur avec ce qui est représenté s'établit.

2. Gallagher (2008) les appelle aussi « représentations minimales ».

Ces nouvelles approches, qui semblent réconcilier le pouvoir explicatif des représentations et celui du dynamicisme, ouvrent néanmoins la porte à des conséquences théoriques étonnantes. En effet, si les représentations ont des contours flous, et si elles sont intimement dépendantes de l'environnement de l'organisme, peut-on dire qu'elles s'étendent au corps et aux objets qui sont impliqués dans ses interactions? *A priori*, il ne semble pas y avoir de raison particulière de penser le contraire.

Or les représentations ont la réputation d'être, en quelque sorte, la quintessence du cognitif pour la philosophie de l'esprit. Toute la théorie cognitiviste des états mentaux – la théorie des « *beliefs, desires, intentions* » (BDI) – porte sur la relation du sujet à ces représentations. L'intentionnalité – entendue comme étant ce qui permet à la représentation d'être à propos d'autre chose que soi-même – est souvent considérée comme la marque de l'esprit (Crane, 1998; Chisholm, 1957; Brentano, 1973). Même Searle (1983, 1984, 1992), qui fait plutôt de la conscience la marque de l'esprit, considère néanmoins que l'intentionnalité et la conscience sont liées de façon nécessaire, de sorte qu'on ne puisse étudier l'un sans étudier l'autre. Aussi, même si la critique anti-représentationnaliste nous empêche pas de prendre pour acquis qu'il y a un lien nécessaire entre le représentationnel et le cognitif, il n'en reste pas moins que lorsqu'on trouve un système représentationnel – dont les représentations sont *sui generis* (et non dérivées) – on suppose que les représentations dans ce système ont quelque chose de cognitif. Dès lors, si elles s'étendent hors du cerveau, le système cognitif s'étend lui aussi hors du cerveau, au moins dans la même mesure.

Il y a un pas à franchir entre affirmer que les représentations ont des contours flous et qu'elles impliquent l'environnement de l'organisme et affirmer qu'elles sont étendues, mais c'est un pas qu'Andy Clark, Mark Rowlands et Mike Wheeler, les premiers défenseurs de la représentation minimale, semblent être prêts à faire. Ici, il n'est pas banal que les premiers à endosser l'esprit étendu aient quelques années auparavant été les premiers à rénover la représentation. On peut se sentir d'autant plus à l'aise d'affirmer que le débat sur l'esprit étendu a ses racines dans le débat autour des représentations.

1.1.2 L'absence de théorie du cognitif

Le deuxième facteur important dans l'essor de la théorie de l'esprit étendu provient peut-être de l'absence d'une caractérisation suffisamment claire et élaborée de ce qui est cognitif.

Par « caractérisation du cognitif », on entend une heuristique ou une théorie qui nous permet, sur une base justifiable, de discriminer ce qui est cognitif de ce qui ne l'est pas. Autrement dit, une caractérisation du cognitif serait un concept du cognitif qui est communicable – à l'opposé, par exemple, d'une représentation interne – et pour lequel on peut fournir une justification – i.e., qui n'est pas *ad hoc*. Selon cette définition, l'hypothèse de Pylyshyn (1980), qui veut que la cognition soit de la computation et vice versa, serait une caractérisation. En effet, supposant que l'on puisse déterminer ce qui est computation et ce qui ne l'est pas, on aurait une définition communicable, et on pourrait trouver une justification dans la théorie computationnelle de l'esprit. En revanche, un critère pour lequel on ne pourrait trouver de justification ne serait pas non plus une caractérisation, ne serait-ce que parce que la justification est ce qui lie le critère à la théorie scientifique et philosophique pertinente.

Quelques remarques peuvent être faites sur les conditions que nous émettons pour déterminer ce qui est une caractérisation et ce qui ne l'est pas.

D'une part, la condition de discrimination semble exclure d'emblée une description de l'esprit qui serait plutôt axé son aspect phénoménal, c'est-à-dire qui servirait moins à discriminer le cognitif du non-cognitif qu'à caractériser le phénomène tel qu'il apparaît à la première personne. On pourrait légitimement arguer qu'un tel projet constitue lui aussi une caractérisation, au sens où il produit une description du cognitif considéré d'un certain point de vue, laquelle peut s'avérer assez exhaustive. Cependant, comme notre question portera sur l'extension et la localisation de l'esprit, une telle description ne conviendra que dans la mesure où elle produit de l'information qui nous permet d'y répondre – et donc, qui nous permet de trancher sur le statut cognitif des objets physiques.

D'autre part, l'usage de la dénomination « caractérisation », par opposition à « critère » par exemple, se veut une façon de s'éloigner d'un certain prototype du critère qui est défini à l'aide de conditions nécessaires et/ou suffisantes. Comme on le verra, un des soucis de ce mémoire sera de neutraliser certains biais découlant de l'usage d'outils conceptuels très usités et donc plus familiers, mais pas nécessairement plus appropriés à l'usage qu'on en fait. Pour éviter les connotations que traîne avec elle la notion de « critère » quant aux moyens par lesquels on produit une caractérisation, il convient donc d'employer un mot moins chargé.

Le besoin d'une caractérisation de l'esprit s'exprime différemment en sciences cognitives et en philosophie de l'esprit. En sciences cognitives, on parle plutôt de « système cognitif », et il est, en quelque sorte, l'*explanandum* commun qui rassemble les différentes disciplines qui composent ce champ d'interdisciplinarités. En philosophie, le concept d'esprit se fait sur le fond d'un fossé théorique entre monde et esprit, de sorte que celui-ci doit être capable d'expliquer – ou de réduire – le phénomène. Dans un cas comme dans l'autre, la recherche de caractérisation découle d'enjeux fondamentaux. Cependant, comme il s'agit de besoins différents, les stratégies utilisées sont aussi différentes.

Sciences cognitives

En sciences cognitives, malgré l'importance du concept de « système cognitif », et malgré le fait que cette expression et ses dérivés (« cognition », « cognitif ») soient très usités, la communauté scientifique n'a pour le terme « cognitif » aucune définition qui fasse consensus.

Cette situation ne semble pas entraver la pratique de la recherche outre mesure, où une définition intuitive d'usage remplace la définition formelle. De fait, l'essentiel de la recherche ne se situe pas au niveau des concepts très généraux, comme celui du cognitif : les chercheur·ses ont plutôt l'habitude de se donner un problème spécifique, portant sur tel ou tel type de mécanisme ou fonction. L'intérêt de se

doter d'une caractérisation du cognitif lorsqu'on étudie le fonctionnement d'une synapse, un type de réseaux de neurones ou le développement de la théorie de l'esprit chez les enfants existe : le concept de cognitif détermine l'horizon de la recherche, et procure un contexte au sein duquel on peut situer les questions de recherches particulières et les études qui tentent d'y répondre. Cependant, pour ce faire, on peut le plus souvent se contenter d'un concept d'usage du cognitif, qui en donne une approximation : si un changement dans la caractérisation précise du concept de cognitif peut avoir un impact, par exemple, sur la portée d'une étude, elle n'affectera probablement pas les hypothèses centrales qui y sont adressées. Aussi, s'il est découvert que la conception que l'on avait du cognitif était mauvaise, la recherche empirique restera valable, du moins en ce qui concerne ses acquis les plus importants.

Cependant, les sciences cognitives ont tout de même pour objectif d'expliquer le cognitif – c'est cet objectif qui motive que des disciplines aussi différentes que la psychologie, la neurologie, la philosophie et l'informatique se rassemblent sous un même projet de recherche. Cette question est donc centrale, et le fait de ne pas avoir su ne serait-ce que la circonscrire reste toujours un peu vexant.

Plus substantiellement, l'absence de caractérisation de l'esprit fait qu'il n'y a aucune façon de savoir si les programmes de recherche des sciences cognitives et les institutions qui les font avancer sont fondés sur de bonnes bases théoriques. Même si deux cadres explicatifs sont capables d'expliquer un même phénomène, ils ne sont pas pour autant équivalents, car un cadre explicatif peut nous amener à nous concentrer sur un phénomène plutôt qu'un autre, et à inspirer une direction plutôt qu'une autre. Comme le montre Lakatos (1970), les programmes de recherche scientifique sont fondés sur un ensemble d'énoncés (implicites ou non) qui détermine le développement des hypothèses et des théories. Si, cependant, les fondements théoriques qui sont à la base de ces programmes de recherche ne sont pas explicites, il devient difficile d'évaluer le programme de recherche.

Bref, bien que le concept du cognitif soit un concept central en sciences cognitives, l'impétus derrière une caractérisation du cognitif ne vient peut-être pas vraiment

d'un besoin pressant, mais plutôt du malaise funeste qu'on peut avoir lorsqu'on agit à l'aveuglette.

Les sciences cognitives comportent leur lot de caractérisations. La plus fameuse est certainement celle de Pylyshyn (1980), mentionnée plus haut. En ce qui concerne la computation, le débat est encore ouvert à savoir si le système cognitif central humain est mieux compris par un modèle entièrement computationnel (e.g. Fodor, 2001; Schneider, 2011) ; mais la possibilité qu'une partie importante de la cognition humaine puisse ne pas l'être suppose que la cognition n'est pas nécessairement computationnelle. Par ailleurs, on peut, comme Van Gelder (1995) le montrait, appliquer une description computationnelle du gouverneur de Watt, mais dirons-nous pour autant qu'il est computationnel ?

Von Eckardt (1995) a fait une tentative beaucoup plus soutenue pour produire une caractérisation. Cherchant à décrire l'unité sous-jacente aux sciences cognitives, elle souligne qu'à ce stade de leur développement, les catégories fondamentales des sciences cognitives ne sauraient s'exprimer qu'à l'aide d'un vocabulaire de sens commun (1995, 47), ce qui semble forcer une « caractérisation » un peu floue³. Ainsi, la liste de propriétés que Von Eckardt utilise pour caractériser les capacités cognitives, par exemple, ne se lit pas comme un ensemble de conditions strictement nécessaires, mais comme un ensemble de repères. Les propriétés elles-mêmes sont sujettes à interprétation, mais, toutes ensemble, elles dirigent nos intuitions en la matière et suggèrent une certaine idée du cognitif.

Cependant, Thagard (1995) déplore que Von Eckardt laisse de côté plusieurs aspects des sciences cognitives (dynamicisme, phénoménologie, intégration des données neurologiques, etc.), et s'intéresse démesurément à l'opinion des cognitivistes. Le concept du cognitif qu'elle obtient n'est donc pas tant un concept du cognitif en général qu'une caractérisation d'un concept fondamental du projet cognitiviste.

On peut se demander si ce problème ne peut être réglé en gardant l'approche

3. Elle prend la forme d'un *cluster concept* (Searle, 1958; Devitt et Sterelny, 1999), notion sur laquelle nous reviendrons en section 1.5.

de Van Eckardt, c'est-à-dire, *grosso modo*, l'induction d'un concept à partir du consensus d'un groupe d'experts, et en étendant le groupe d'experts à l'ensemble des domaines de la cognition. Cependant, l'avantage de travailler avec un groupe partageant des bases théoriques communes est non seulement qu'il y a des concepts fondamentaux communs, mais surtout que le travail scientifique qui s'y fait peut être ramené à un même vocabulaire et une même ontologie. Il n'y a aucune garantie que l'on puisse trouver un consensus sur un concept qui s'exprime sur deux ontologies différentes, et en tout cas, le travail pour produire ce consensus devra être colossal. De plus, même en ayant collecté un ensemble de propriétés communes, encore faut-il isoler celles qui sont pertinentes et justifier ce choix. Pourquoi ne pas travailler directement sur les principes qui nous permettraient d'effectuer cette discrimination? Bref, s'il n'est pas impossible de suivre cette suggestion et de travailler à partir des bases jetées par Von Eckardt, on peut se demander si le travail en vaut la chandelle.

Philosophie de l'esprit

Si les sciences cognitives échouent à produire une caractérisation du cognitif, on pourrait espérer que la philosophie de l'esprit puisse produire une caractérisation de l'esprit. D'une part, c'est le fait de la philosophie de chercher à clarifier et définir les concepts fondamentaux comme celui-ci. D'autre part, l'identité de la philosophie de l'esprit est fondée sur le fossé entre l'esprit et le monde, et une grande partie de sa mission est de décrire et d'expliquer ce que l'esprit a de particulier par rapport au monde. Aussi, alors que l'importance de la caractérisation du cognitif en sciences cognitives vient de ce qu'elle peut orienter la recherche dans la bonne direction, son importance en philosophie de l'esprit vient de ce qu'elle peut rendre compte du caractère particulier de l'esprit.

L'opinion dominante en philosophie de l'esprit (Dennett, 1989; Jacob, 2010) fait souvent remonter ce particularisme à Quine (1960), même s'il n'est pas sans précurseurs. Celui-ci notait en effet que le « vocabulaire intentionnel » associé au

domaine mental était irréductible au vocabulaire non-intentionnel, en ce sens que les concepts du vocabulaire intentionnel sont sous-déterminés par le comportement. Autrement dit, l'information renvoyée par le comportement d'un-e agent-e est toujours insuffisante pour lui attribuer une attitude comme un désir ou une croyance. Quine rejoignait ainsi les conclusions de Chisholm (1957) et de Brentano (1973), mais plutôt que d'y voir, comme eux, un signe que le langage intentionnel est indispensable et qu'une science autonome doit l'étudier, il y voyait plutôt une preuve de sa vacuité.

Curieusement, cet état de choses ne l'empêchait pas de concéder que, si le langage intentionnel n'était pas approprié pour découvrir « *the true and ultimate structure of reality* », il pouvait l'être dans la pratique quotidienne. On peut certes se demander comment il est possible qu'un vocabulaire évoquant des concepts vides puisse avoir du succès dans le domaine pratique. De fait, le jugement de Quine sur la vacuité du langage intentionnel ne s'avère pas sans appel. Davidson (1963), s'appuyant sur une idée d'Anscombe (1957), décrit le langage intentionnel comme fonctionnant sur la base de ses propres présuppositions (la rationalité des agent-es) : une description, par exemple, d'un état mental (e.g. Hortense croit que le canard est trop cuit) dans un langage intentionnel est alors vraie et sensée *sous une certaine description*. Ainsi, si langage intentionnel et non-intentionnel sont indépendants, ce n'est pas parce qu'ils sont métaphysiquement indépendants, mais parce qu'ils le sont ontologiquement. Ils ne découpent pas le monde de la même façon – si jamais il y a un plus petit dénominateur commun entre les deux façons de décrire, il nous est inconnu. Davidson reste donc attaché au monisme physicaliste, mais inaugure en revanche un dualisme des descriptions.

Dès lors, ce qui était chez Quine une fâcheuse ambiguïté quant à la valeur du langage intentionnel se cimente chez Davidson et chez ses successeurs en une tension dans la compréhension de l'esprit. En effet, de ces épisodes la philosophie de l'esprit retient, à divers degrés et sous différentes formes, à la fois le monisme physicaliste, le dualisme des descriptions et l'autonomie du langage intentionnel. Le mental est donc particulier – représenter et expliquer cette particularité, ou la possibilité de cette particularité, devient donc un enjeu fondamental de la philo-

sophie de l'esprit. Une des contraintes qui s'y exprime est la nécessité de rendre ce que le mental a de particulier tout en le présentant à l'intérieur du monde physique. Autrement dit, les attributs du mental que la philosophie de l'esprit doit trouver sont aussi ceux qui sont pertinents pour le projet de naturalisation de l'esprit.

Quelles sont ces particularités qui font l'essence de l'esprit ? En voici quelques-unes :

1. *Intentionnalité* : On appelle « intentionnalité » cette caractéristique qu'ont certains ou tous les états mentaux qui est d'être à propos d'autre chose que soi-même. Ainsi, l'intentionnalité est une condition nécessaire à la formation des représentations, lesquelles forment le contenu des attitudes propositionnelles. Elle est aussi cette caractéristique qui fait que la conscience est souvent (ou toujours, dépendant du philosophe) conscience de quelque chose – bref, elle permet à l'organisme de se représenter et d'être conscient du monde. Il s'agit donc d'un élément fondamental de la vie cognitive.

Il est utile pour la suite de mentionner ici que l'intentionnalité se dit différemment s'il s'agit d'un phénomène mental ou d'une représentation externe, par exemple un mot écrit sur une feuille de papier. Dans le second cas, on dira que la force intentionnelle du mot vient de l'interprétation de celui ou celle qui le lit – on dira donc qu'il possède de l'intentionnalité dérivée. L'interprète, en revanche, est capable de produire de l'intentionnalité non-dérivée, dont le pouvoir intentionnel n'est dû à aucun autre interprète. Aussi seule l'intentionnalité non-dérivée est-elle une caractéristique propre à l'esprit.

2. *Conscience* : Par conscience, on doit comprendre quelque chose comme « res-senti » ou « expérience phénoménale », et non pas une quelconque opération cognitive ou quelque chose qui peut se réduire à un état neurologique. La conscience est souvent considérée comme étant le *relatum* qui est mis en relation avec l'objet dans la relation intentionnelle (Searle, 1992). Cependant, même lorsqu'on ne se prononce pas sur son rôle dans les opérations ou phénomènes cognitifs, elle a une place de choix de par ceci qu'elle définit le domaine du phénoménal – lequel, dans l'optique de la philosophie de l'esprit, reste souvent identifié au domaine de l'esprit, vu d'une certaine perspective.

3. *Causalité anormale* : Un autre de ces attributs est la façon particulière que la causalité a de s'exprimer dans le domaine mental : ainsi, Davidson (1970) notera l'anomalisme du mental, soit l'absence de loi liant les états mentaux et l'action, et ce même si les uns sont utilisés pour expliquer les autres. I.e. on peut expliquer que Rodrigue ait éteint la télé à cause de sa croyance que son programme est ennuyeux, mais comme cette croyance n'amène pas toujours la même réponse, il n'y a pas de loi liant la croyance au fait d'éteindre la télé (sans quoi Rodrigue eût éteint la télé un quart d'heure plus tôt).
4. *Accès à la première personne* : Un autre phénomène concerne l'accès privilégié de la première personne à ses états mentaux. On pourrait dire qu'on a, en tant qu'individu, un accès exclusif à certains objets – e.g. le ressenti des couleurs, la douleur, nos pensées intimes – alors que d'autres sont d'accès public – e.g. mon bureau, le bordeaux en dégustation, la nation québécoise, la théorie de la relativité générale.

Plusieurs tentatives ont été faites pour caractériser l'esprit en terme de caractéristiques particulièrement mentales. Par exemple, il y a eu plusieurs tentatives d'établir une « marque du mental » : quelque chose qui nous permettrait de dire si quelque chose est mental ou ne l'est pas. L'intentionnalité est certainement la plus populaire (Brentano, 1973; Crane, 1998; Adams et Aizawa, 2008)⁴, la conscience suit (McGinn, 1995), puis viennent d'autres critères comme l'anomalisme (Davidson, 1970) et l'incorrigibilité Rorty (1970b). Pour les marques du mental comme pour les tentatives de caractérisation plus élaborées, la stratégie est généralement la même : prendre une caractéristique – telle qu'elle apparaît dans la description en langage intentionnel – et en faire un critère pour que quelque chose soit considéré esprit.

Cependant, le genre de soucis que l'on avait avec la computation peut aussi s'appliquer sur la description intentionnelle, dans laquelle s'expriment les différentes marques du mental. Le computationnalisme étant enfant du cognitivisme, la description computationnelle que Van Gelder appliquait au gouverneur de Watt emploie des termes intentionnels. Si l'on est capable de faire coller une telle descrip-

4. Adams et Aizawa (2008) parlent de « contenu non-dérivé », mais c'est tout à fait analogue au concept d'intentionnalité non-dérivée.

tion au gouverneur de Watt, on peut forcément lui attribuer des représentations et donc de l'intentionnalité. Si intentionnalité il y a, alors il lui faut un *relatum* – pourquoi pas une conscience ? – et donc aussi un accès privilégié, et toute la faune de la description intentionnelle peut suivre. Bref, il y a peut-être moyen de défendre l'indéfendable – à savoir que le gouverneur de Watt est un bon candidat pour posséder un esprit.

La parade à ce raisonnement et à sa conséquence contre-intuitive est de faire la distinction entre l'intentionnalité non-dérivée, que possèdent les êtres humains, et l'intentionnalité dérivée qui découle de notre interprétation du monde. La description intentionnelle du gouverneur de Watt serait due à notre interprétation, de sorte que celui-ci n'est pas intentionnel au même sens où l'est notre esprit. Si on peut distinguer entre intentionnalité dérivée et intentionnalité non-dérivée, on a peut-être alors un bon repère pour distinguer le mental du non-mental.

Cependant, pour appliquer cette distinction, il n'y a pas de « marque du non-dérivé » : on conçoit intuitivement que l'intentionnalité qui se forme dans ma conscience et celle que j'attribue à un symbole sur une pancarte sont de nature différentes, parce que la pancarte n'aurait pas le sens qu'elle a si on la catapultait dans l'espace, où aucune forme de vie n'est en mesure de la lire ; mais on ne sait pas en vertu de quoi mes représentations internes sont non-dérivées. Le problème se complique donc lorsque l'on parle d'un organisme tiers : si je lui attribue des représentations, est-ce seulement le fait de mon interprétation, ou est-il capable de les produire lui-même ?

On peut encore déplacer le problème vers une intuition plus générale : de fait, on possède une capacité intuitive à attribuer des états mentaux aux autres, et cette capacité ne nous sert pas si mal au quotidien. Il semble certes que les humains (et d'autres animaux, à divers degrés) disposent d'un robuste instinct pour attribuer esprit et états mentaux à nos semblables et à d'autres créatures – ce que les psychologues appellent notre « théorie de l'esprit ».

Pourtant, si on considère la seule tâche de détecter un organisme possédant une

certaine intelligence propre (et donc susceptible de posséder des représentations non-dérivées), on constate que l'intuition n'est pas particulièrement efficace. En effet, l'ensemble des entités auxquelles on est porté à attribuer des états mentaux peut varier considérablement, autant par défaut que par excès. Ainsi, d'un côté, on est tenté d'attribuer des états mentaux à des triangles qui se meuvent sur un écran (Heider et Simmel, 1944), et les enfants attribuent volontiers des croyances et un statut moral à des jouets, mais de l'autre, Descartes faisait des animaux des automates sans esprits, et Malebranche leur niait en prime toute intelligence. De telles variations montrent bien que l'intuition ne peut suffire à elle seule : il faut un critère. En ce sens, les caractérisations en termes de caractéristiques mentales ne semblent que déplacer la question.

Dans le même esprit, on peut se demander si les caractéristiques propres à l'esprit ne sont pas en fait des artefacts de la description. C'est certainement le cas pour la causalité anormale : si avoir une raison pour faire x ne mène pas à x , il y a probablement des facteurs au niveau des processus qui réalisent le fonctionnement cognitif de l'agent·e qui l'ont empêché – i.e. quelque chose dans le cerveau de Rodrigue a agi d'une telle façon que, malgré son désir de se lever, il est resté assis devant la télé. L'absence de loi ne s'explique que parce que le niveau de description fait abstraction de facteurs importants dans l'explication causale.

Peut-on en dire autant des autres caractéristiques ? En fait, même si la conscience, l'intentionnalité et l'accès privilégié représentent des propriétés réelles, indépendantes de la description, la description intentionnelle nous amène certainement à les décrire à la lumière des hypothèses de la méthode.

Concernant l'incorrigibilité, Rorty lui-même concède qu'elle suit d'une notion de conscience immédiate, par opposition à un monde distal :

The mind-body problem is an offspring of the theory that knowledge consists in the having of certain representations of reality (including perceptual ones), by the subject. As [Wallace] Matson has recently pointed out [in “Why Isn't the Mind-Body Problem Ancient?”], the Greeks had a soul-body problem but not a mind-body problem—or,

at least, not the mind-body problem which has bothered philosophers from Descartes to Feigl. Before *this* mind-body problem can be made to seem urgent (as Matson also notes) one has to have the notion of ‘immediate awareness’, and to believe that the things we want to know about (tables, other men, stars, the moral law, and the gods) are not things which we are immediately aware of. Once one believes all this, one will have to grant the existence of a realm to contain the objects of immediate awareness. This will be the Mind, or the Subject *qua* Subject. Psychophysical dualism follows from epistemological dualism. In the great systems of the Cartesian period, the primary task of ontology was to get the Subject and the Object back together. Solutions to the mind-body problem appear as simple corollaries to solutions of the Problem of Knowledge. (Rorty, 1970a, 278)

Cependant, le problème se pose autrement si, au lieu d’employer le concept traditionnel de représentation, on emploie des représentations incarnées (cf. *infra* p. 5) : rien ne leur empêche d’être l’objet d’une conscience immédiate, mais le fait qu’elles le soient ou non est indépendant de leur valeur épistémologique. La représentation incarnée est adéquate dans la mesure où elle donne à l’agent·e les informations nécessaires pour l’action ; dans la mesure où les diverses fonctions dans lesquelles elle est impliquée sont remplies adéquatement. À la base, le problème épistémologique est donc un problème d’apprentissage machine : il n’y a donc besoin ni de conscience immédiate, ni d’établir un lien entre la représentation et son objet qui dépasse le contact de l’agent avec l’objet à travers l’action.

Or, lorsqu’on cesse de considérer qu’on a besoin de conscience immédiate pour avoir des représentations, les perceptions conscientes ne jouissent d’un accès privilégié que dans un sens très restreint : certes, seule la chauve-souris sait ce que c’est que le ressenti d’une chauve-souris, mais l’agent·e n’est plus nécessairement le meilleur témoin de ses croyances. Par exemple, on sait qu’un·e agent·e ne réussit pas toujours à rendre compte des biais qui l’affectent (e.g. Pronin, 2007). Ainsi, on observe que lors de processus d’embauche, même les femmes peuvent être moins portées, à compétences égales, à embaucher d’autres femmes (e.g. Moss-Racusin *et al.*, 2012). Or on ne voudra pas nier que ce biais – dont l’agent·e serait souvent incapable de rendre compte – appartient bien au système cognitif. Aussi, dans un cadre de représentation minimaliste, l’accès privilégié ne vaut que pour le ressenti,

pas pour le contenu.

Le même genre de raisonnement prévaut pour la conscience : celle-ci ayant perdu son lien intime avec la représentation, il n’y a plus moyen de l’étudier à travers celui-ci comme le proposait Searle (1984). C’est pourquoi elle devient l’objet du « problème dur » (Chalmers, 1995) : séparée de l’intentionnalité ou d’autres caractéristiques de l’esprit, la conscience n’est pas définissable fonctionnellement. Une définition fonctionnelle nous ouvre à la possibilité d’une explication à travers des mécanismes qui remplissent la fonction, mais comme la conscience n’en dispose plus, il devient très difficile d’imaginer comment on peut déterminer si un mécanisme réalise ou non une conscience.

Quant à l’intentionnalité, dans sa nouvelle incarnation « située » (Millikan, 1984, 2000; Menary, 2009, 2007), elle reste une caractéristique mentale par excellence – mais dans une version modifiée, à l’image des nouvelles théories de la représentation. Dans cette nouvelle incarnation, elle continue dans certaines théories (Menary, 2007) à jouer un rôle de caractéristique de l’esprit. Cependant, suite à son divorce avec la conscience, elle ne peut plus jouer le rôle de « marque du mental », car il n’est plus possible de l’associer, comme le faisaient Brentano et Crane, à toute expérience phénoménale. Aussi verra-t-on l’intentionnalité se manifester plus libéralement, jusqu’à la faune unicellulaire et à des robots ou des simulations informatiques un peu rudimentaires, voire à des systèmes encore plus frustes, pour lesquels il est plutôt contre-intuitif d’affirmer qu’ils sont des systèmes cognitifs.⁵

En bref, les caractérisations de l’esprit se situent à un niveau de description abstrait, ou alors sont intimement liées à l’expérience phénoménale, ce qui les rend peu utiles pour informer notre connaissance de la réalisation matérielle de l’esprit, qui se situe à un niveau plus concret. En conséquence, elles ne permettent pas de déterminer si un mécanisme ou un processus est cognitif ou pas.

5. Notons ici qu’entre l’externalisme du contenu, associé à Putnam (1973), et l’externalisme véhiculaire (expression que l’on emploie parfois comme synonyme de « thèse de l’esprit étendu » (e.g. Hurley, 2010), seul le second peut répondre à la question de l’extension de l’esprit. En effet, le contenu appartient au vocabulaire intentionnel ; en ce sens, il ne se définit pas en termes de mécanismes, et toute tentative de le traduire en ces termes se heurte aux difficultés soulevées plus haut pour les caractéristiques de l’esprit appartenant au niveau intentionnel.

1.2 La question

1.2.1 Caractérisation et extension

Les caractérisations de l'esprit issues d'une philosophie de l'esprit fondée sur une formulation dualiste du problème corps-esprit ont montré leurs limites, et les sciences cognitives ne possèdent pas de caractérisation générale du cognitif qui vaille autant, par exemple, pour une approche cognitiviste que pour une approche dynamiciste. Les disciplines de l'esprit semblent donc incapables de formuler le concept qui les rassemble. Bien que ce fait ne soit pas nécessairement au centre de l'attention de la philosophie de l'esprit et des sciences cognitives, la constatation est là (Von Eckardt, 1995; Menary, 2010; Aizawa, 2010; Golonko, 2010; Barker, 2010).

Ken Aizawa⁶ voit en cette situation une condition habilitante (*enabling condition*) à l'esprit étendu. Cependant, l'échec de certains projets de caractérisation ne devrait pas nécessairement mener à une hypothèse aussi radicale. Il y a donc d'autres conditions qui expliquent que la question de la caractérisation de l'esprit puisse, au moins dans l'esprit de certains philosophes, se transformer en une question de localisation et de détermination de son extension.

D'une part, il y a la question du niveau d'analyse. À travers son histoire, l'esprit en est venu à se situer sur différents niveaux d'analyses, avec différentes ontologies et différents fonds métaphysiques. Au niveau métaphysique, à une extrême, Descartes a placé l'esprit dans la *res cogitans*, une essence indépendante du monde physique. À l'autre extrême, les behavioristes ont voulu limiter l'étude de l'esprit à celle du comportement. À cela, on peut ajouter les gens qui, comme Brentano, ont situé l'esprit au niveau de l'expérience phénoménale. Aujourd'hui, cognitivistes, phénoménologues et autres s'entendent généralement pour dire que le mécanisme qui cause ce comportement est le domaine métaphysique où se réalise l'esprit. Chacune de ces métaphysiques vient avec une ou plusieurs façons de le décrire, souvent avec son propre vocabulaire et sa propre ontologie. Autrement dit, il n'y

6. Dans un commentaire à Golonko (2010).

a pas pour l'esprit de catégories fondamentales qui s'imposent spontanément.

Cela n'empêche pas certains de croire que ces différentes catégories doivent se recouper quelque part. Ainsi, se manifeste dans certaines théories, comme le computationnalisme, la supposition que les mécanismes de la pensée, tels qu'ils apparaissent dans une description intentionnelle, sont à peu près isomorphes aux mécanismes qui la réalisent. Comme le dit Fodor (1990, 20) : « ... we have no philosophically impressive reason to doubt that [Sherlock] Holmes's train of thoughts went pretty much the way that he says it did. ».

Cette idée a pourtant perdu le lustre qu'elle a eu pour les cognitivistes. D'une part, comme le défend Wimsatt (2007, chp. 9), on ne peut s'attendre dans un système biologique à ce que le découpage ontologique d'un niveau d'analyse soit systématiquement repris au niveau inférieur, où on ne ferait que lui ajouter un découpage supplémentaire – il appelle cette absence d'alignement des découpages ontologiques « *complexité descriptive* » – et ce notamment parce que les interactions dans ces systèmes n'ont pas à suivre les frontières, par exemple, d'un organe ou d'une cellule. Or le système cognitif – connu pour sa plasticité, et sa capacité à connecter toutes sortes d'informations qui proviennent d'un peu partout – est peut-être le système qui a les meilleures chances d'exhiber de la complexité descriptive. En ce sens, l'argument que donne Fodor – à savoir que, dans le fonctionnement de l'esprit, les raisons ont le potentiel d'être des causes – semble insuffisant.

D'autre part, les prédictions empiriques qui se déduisent de la thèse de l'isomorphie ne semblent pas passer le test de l'expérience. Prenons par exemple le concept de représentation : l'usage qu'on en fait dans le langage ordinaire suggère que la représentation doit être fortement découplable de son contexte ; or des soucis comme le problème de l'ancrage ont forcé les philosophes à reconsidérer cette position. De même, rien dans les descriptions intentionnelles de nos raisonnements ne laisse soupçonner que l'on recrute des ressources externes pour accomplir des tâches cognitives ; or il semble que ce soit plutôt fréquent (e.g. Kirsh et Maglio, 1994; Goldin-Meadow, 2005). Enfin, comme le note Hurley (1998, 5), citant des

études séminales dans l'application de la théorie des systèmes dynamiques :

An essential lesson we learn from studying nonlinear systems is that discontinuities and structure at a higher level can emerge unpredictably (though deterministically), with no isomorphism at the underlying level.

Si les phénomènes intentionnels ne nous donnent pas l'information pertinente pour isoler les mécanismes de la cognition, alors l'esprit ne peut se caractériser au niveau intentionnel. Ce n'est pas dire qu'on ne peut prendre en compte des contraintes venant du niveau intentionnel, mais l'absence d'isomorphie fait en sorte qu'il est périlleux de partir d'une description intentionnelle et de tenter de l'implémenter au niveau du mécanisme. Bref, s'il faut caractériser l'esprit, il faut le faire au niveau du mécanisme.

C'est ce qui nous amène à l'extension de l'esprit. Bien que le nom suggère une région géométrique où seraient contenus les mécanismes cognitifs, la question porte, en pratique, sur les caractéristiques qui nous permettent de déterminer cette région – type de tissus organiques, patrons d'interactions, etc. Déterminer l'extension de l'esprit, c'est donc déterminer la caractérisation de l'esprit au niveau du mécanisme. En ce sens, la question qui nous préoccupe est une nouvelle incarnation de la question de la caractérisation de l'esprit, adaptée à un certain contexte scientifique. Cependant, c'est aussi une étape vers une compréhension plus complète de l'esprit : une fois que l'on connaît l'extension de l'esprit, on peut l'investiguer et en tirer des connaissances pertinentes. La détermination de cette extension est donc à la fois une caractérisation de l'esprit ou du cognitif (orientée vers un objectif spécifique qui est la détermination de l'extension) et une étape vers une caractérisation plus complète.

En conséquence, étant donné un contexte scientifique comme celui qui est d'actualité dans les sciences cognitives et qui exige d'une caractérisation qu'elle nous permette de discriminer dans l'espace physique les phénomènes qui sont cognitifs de ceux qui ne le sont pas, la question de l'extension en vient à se fondre à la question de la caractérisation, pour ne constituer qu'une seule question. D'une part, parce que, dans les faits, la question de la caractérisation, lorsqu'elle vient

avec la contrainte d'être capable de déterminer quel genre de mécanisme réalise l'esprit, en vient à demander explicitement une extension ; et d'autre part parce que la question de l'extension ne peut trouver de réponse sans rechercher des caractéristiques de l'esprit qui permettent de marquer son territoire, et parce que la raison d'être de cette question est finalement de contribuer à la caractérisation de l'esprit. En ce sens, en notre temps et en notre contexte intellectuel, il faut considérer la question de l'extension dans la perspective de la question de la caractérisation, et inversement.

1.2.2 Problématique

Dans le cadre de ce mémoire, ces questions ne sont appelées qu'à jouer un rôle d'arrière-plan : elles forment le contexte dans lequel s'inscrira notre problématique. Cependant, qu'elles soient aussi intimement liées est assez important pour l'ensemble du projet – notamment parce que cela suggère que l'une des deux questions ne saurait être l'échappatoire de l'autre. Autrement dit, on ne saurait dire qu'il faut reporter notre examen de la question de la caractérisation au jour où on aura une idée claire et bien circonscrite des mécanismes qui réalisent la cognition, comme on ne peut non plus, en cherchant les frontières de l'esprit, en appeler à un retour à une théorie ou à un critère de l'esprit (comme le font Wheeler, à paraître; Rowlands, 2010b; Barker, 2010). Renvoyer à la prochaine question, dans ce cas, relève d'une certaine myopie – ou alors faudrait-il justifier en quoi ce renvoi nous fait avancer, et ne mènera pas à ce qu'on revienne à nouveau à la question que l'on voulait quitter.

Par ailleurs, le fait de choisir une question plutôt que de les aborder les deux de front ne peut qu'affecter la façon dont on entend y répondre, puisqu'une telle perspective affecte l'horizon des heuristiques qui se présentent à nous. Si l'on cherche simplement une description de l'esprit, et que l'extension ou la localisation importent peu, alors on voudra probablement qu'elle s'accorde avec le principe fonctionnaliste de réalisabilité multiple, qui veut qu'un même processus

ou qu'une même fonction cognitive puisse être implémenté dans divers médias. Aussi, pour que son modèle soit pertinent pour un maximum de programmes de recherche, semblera-t-il *a priori* avantageux de ne pas employer d'information provenant d'une approche particulière traitant de l'implémentation des processus et fonctions cognitives, et donc, à toute fin pratique, d'ignorer la recherche en ce domaine. Inversement, si l'on cherche l'extension, un souci similaire de neutralité nous amène-t-il, en regard à la multiplicité des formes que pourraient prendre les processus et fonctions cognitives, à désirer un critère qui peut toutes les accepter.

Or, qui trop embrasse mal étroit. Il n'y a aucune raison, en l'absence d'indices laissant croire le contraire, de penser que l'information sur la forme, l'implémentation ou la localisation de ce qui constitue le cognitif ne soit pas pertinent à la formation du concept de cognitif ou d'esprit. On peut généraliser ce principe, et l'énoncer ainsi :

Principe d'exhaustivité. Pour toute question, une partie d'un domaine de recherche ne saurait être écartée s'il n'y pas de bonne raison suggérant explicitement que des informations pertinentes ne sauraient s'y trouver.

Il faut ici interpréter ce principe comme un appel à exercer sa prudence envers des raccourcis de la pensée, plutôt que comme un appel à évacuer complètement ces derniers du processus. Il est naturel et sain d'adopter des heuristiques lorsque l'on est confronté à un problème (Gigerenzer et Brighton, 2009), et même en sciences, la pratique est de mise (Dunbar, 1997). Cependant, il n'en demeure pas moins qu'il faille, au moins après coup, revisiter les raccourcis qu'impliquent ces heuristiques pour traquer les biais où ils peuvent nous induire. C'est le genre d'ouvrage qu'entreprendra ce mémoire : revisiter le travail des philosophes qui ont répondu à la question de l'extension pour déterminer s'il ne cache pas des erreurs ou des omissions.

Le cas particulier qui nous intéresse ici est le concept de frontières dans la littérature sur l'esprit étendu. Celles-ci, en effet, constituent une façon peu coûteuse en termes d'information de départager le substrat matériel réalise le cognitif (ou est en mesure de le réaliser) de celui qui ne le réalise pas. C'est particulièrement

vrai dans le cas de la position de l'internaliste véhiculaire : si les frontières du cognitif sont celles du cerveau, alors le tracé est déjà fait. Pour l'externaliste, cependant, formuler le domaine du cognitif en terme de frontière est aussi une stratégie économique. D'une part, comme l'esprit étendu est d'abord une position critique de l'internalisme, on voudra d'abord assurer sa position en se concentrant sur un ou plusieurs points de contention. Conserver le concept de frontière et l'employer pour représenter le différent théorique permet de mesurer la distance entre la théorie critiquée et la théorie avancée comme l'alternative. La ou le partisan-e de l'esprit étendu n'a alors qu'un front bien défini à défendre, et il ou elle peut déterminer si sa position a bien été argumentée ou non. D'autre part, définir une extension en déterminant sa frontière est économique en termes d'information. Il est beaucoup plus ardu, par exemple, de définir des régions multiples susceptibles de faire partie du système cognitif d'un organisme, ou de définir des aires floues, à diverses probabilités ou degrés d'appartenance.

Dans la mesure où elle représente une façon commode et rapide de représenter sommairement leurs positions, les différents partis dans le débat sur l'extension de l'esprit sont justifiés d'employer le recours aux frontières comme heuristique pour répondre à la question de l'extension. Cependant, vient un temps où les lignes du débat s'étant établies, on peut revisiter les concepts qui le sous-tendent.

En l'occurrence, le concept de frontière n'est pas la seule alternative disponible – comme on entend le montrer un peu plus loin, en section 1.5. On aurait pu par exemple, déterminer que l'esprit se réalise à divers niveaux sur plusieurs substrats, de sorte que les contours soient flous au point de s'effacer – un peu comme pour ce qu'on voit avec les contours d'un écosystème (Degré, 2011). Le travail cognitif pourrait aussi se réaliser sur divers médiums disjoints – de la même façon que plusieurs individus contribuent à produire une idée à travers une conversation, ou de la même façon que des mécanismes indépendants les uns des autres dans les robots de Brooks (1986) contribuent à ce qui semble être d'un point de vue externe le comportement d'un agent robotique unique.

Lorsque la question n'était que de savoir si l'esprit limitait son extension aux

tissus nerveux du cerveau, ou s'il s'étendait au-delà de cet organe, il n'était peut-être pas nécessaire d'être précis quant à l'extension de l'esprit humain, et par conséquent, déterminer celle-ci par une frontière était un luxe dont on pouvait se passer. Cependant, on commence à voir des débats entre externalistes qui portent précisément sur cela (e.g. entre une vision centrée sur l'organisme, comme chez Clark (2008), et une vision qui admet une conception d'un esprit étendu dans un groupe, comme chez Gallagher (2011) ou Hutchins (2011)). Aussi devient-il pertinent de clarifier la forme que doit prendre la cognition si elle est étendue.

Au vu de ces conditions – maturité du discours sur l'extension, présence d'alternatives – mettre en suspens la décision quant à la forme que doit prendre l'extension du cognitif et ouvrir le débat à d'autres façons de la déterminer devrait aller de soi. Cependant, il faut peut-être reconnaître une certaine sagesse dans le conservatisme théorique que l'on retrouve, par exemple, chez Rupert (2009, 18) – qui croit qu'en l'absence de progrès empirique, une nouvelle théorie doit céder le pas à l'ancienne. La confiance que les théoriciens ont implicitement accordée à certaines suppositions qui ne sont jamais remises en question n'est pas nécessairement aveugle : l'intuition des générations précédentes de philosophes tient parfois à des contraintes qui ne sautent pas spontanément aux yeux, mais qui ne redeviennent visibles que lorsque l'on cesse de s'y fier. Il importe donc d'étudier un dogme avant de le rejeter.

Aussi la problématique de ce mémoire sera-t-elle d'étudier le concept de frontières comme façon de répondre à la question de l'extension. Celle-ci appelle à son tour deux sous-questions : d'une part, quelles raisons sous-tendent la présupposition selon laquelle le concept de frontière est approprié pour déterminer l'extension de l'esprit ? Et d'autre part, est-ce que ces raisons sont légitimes, et parviennent-elles à établir que le concept de frontière est bien le concept approprié ?

Notre hypothèse ici est que les alternatives à cette façon d'isoler le lieu physique où se réalise le cognitif sont envisageables. Aussi croyons-nous qu'il n'y a pas, dans le contexte théorique qui sous-tend l'usage du concept de frontière de raison qui légitime son usage à l'exclusion des autres options.

Plus précisément, nous croyons qu'outre la simplicité, cet usage est justifié principalement par ce que nous appellerons dans la suite *l'intuition individualiste*. Celle-ci veut que la cognition soit essentiellement un phénomène individuel, compris comme étant l'apanage d'un organisme distinct, dont les comportements et attitudes cognitives peuvent faire l'objet d'une description de niveau personnel. Dans cette conception, l'individu est l'alpha et l'oméga de la cognition : le principe causal qui lui donne naissance et le principe final en vertu duquel elle œuvre.

L'intuition individualiste permet de faire le pari qu'étant donné que les descriptions du cognitif faites au niveau personnel ramènent tout à l'individu, les mécanismes qui réalisent ces comportements cognitifs doivent se rassembler autour de celui-ci. Par ailleurs, comme la description personnelle ramène de façon non-ambigüe les actions cognitives à l'organisme qui en est la cause, elle permet de faire le pari que les mécanismes qui les réalisent appartiennent aussi de façon claire à celui-ci. En ce sens, elle suggère une extension discrète et délimitable qui se décrit bien à l'aide d'une frontière.

1.3 La démarche

En somme, nous nous trouvons face à trois hypothèses. La principale (H1) est que l'option d'un concept de l'esprit dont la réalisation ne possède pas de frontières déterminées est une option envisageable – c'est d'ailleurs l'hypothèse de ce mémoire. La seconde (H2) veut que, dans le contexte du débat actuel sur l'extension de l'esprit, l'hypothèse inverse – qui affirme que l'esprit doit avoir des frontières – est due à une croyance implicite, l'intuition individualiste. La troisième (H3) est que cette intuition individualiste elle-même est arbitraire, et que rien ne la justifie vraiment.

Ces trois hypothèses ne sont pas indépendantes : si H3 et H2 sont vraies, alors H1 le sera aussi. Aussi notre parcours prend-il exactement cette voie : pour démontrer H1, on s'affaire à démontrer les deux autres hypothèses, ce que nous avons choisi

de faire par une revue des principaux critères du cognitif.

Une supposition que nous avons en adressant H2 est que les rouages logiques qui sous-tendent l'association entre intuition individualiste et frontières de l'esprit sont plus ou moins mystérieux. Autrement dit, il nous semble que les lieux communs explicites de la philosophie de l'esprit et de la philosophie des sciences cognitives ne permettent pas d'établir H2. Eût-ce été le cas, nous aurions pu simplement mettre à jour ces rouages, et la présence de frontière aurait paru découler naturellement de l'intuition individualiste.

Si H2 est vraie, il faut donc que ce soit en vertu de lieux communs implicites – à moins que H2 soit elle-même un lieu commun implicite. Dans tous les cas, la méthode qui semble s'imposer pour établir H2 est l'induction. En effet, on peut établir un échantillon relativement exhaustif comprenant un certain nombre de contextes de discours, et les étudier afin de déterminer si H2 y est vraie. Si c'est le cas pour tous, on peut en induire que H2 est vraie pour le contexte de discours étudié. Certes, l'induction a ses limites ; aussi, si l'on pouvait trouver d'autres lieux communs implicites qui sous-tendent notre hypothèse, la démonstration en serait plus convaincante. Cela représente cependant beaucoup de travail, et on peut penser que le gain qu'on en tirerait ne justifie pas le coût.

On verra dans la suite (chapitre III) que la critique de l'intuition individualiste est déjà largement entamée – notamment par Hurley (2002), dans ce qu'elle appelle la critique du schéma entrée-sortie (*input-output picture*). En ce sens, le gros de la preuve de H3 peut être puisée à même la littérature qui est déjà disponible, et ne pose pas de problème particulier. En revanche, l'intuition individualiste étant implicite, cette critique est souvent limitée par le fait que l'on n'apprécie pas toujours complètement l'étendue de ses effets théoriques, et qu'en conséquence, on ne peut porter la critique à cette extrémité.

En ce sens, pour notre travail, la démonstration de l'hypothèse H2 est particulièrement importante, non seulement parce que celle-ci est la courroie de transmission entre H1 et H3, mais aussi parce qu'elle permet d'étendre la critique de l'intuition

individualiste telle qu'elle a déjà été entamée. Cependant, étant donné le caractère implicite de l'intuition individualiste, cette démonstration est aussi la plus difficile. Aussi, ne faut-il pas se surprendre qu'elle prenne la plus grande partie de ce mémoire.

1.4 Analyse de texte

Que H2 soit la courroie de transmission entre H1 et H3 implique certaines contraintes. Pour que la fausseté de l'intuition individualiste se traduise en la fausseté de la présomption d'une frontière du cognitif, il faut que l'intuition individualiste soit *toujours* la raison nécessaire pour postuler cette frontière. Il faut en effet démontrer que, dans le contexte de discours qui nous intéresse, l'intuition individualiste est une raison nécessaire pour exiger des frontières, et cette hypothèse ne saurait s'accommoder d'un cas où cet impératif se justifie autrement.

Nous avons donc besoin d'une étude qui porte sur l'ensemble des écrits faisant partie de ce contexte de discours. Pour ce faire, il n'est pas question de faire une analyse approfondie de tous les textes et/ou auteur·es qui ont abordé le problème de l'extension de l'esprit : une telle approche serait inutilement longue et ardue. Une méthode appropriée devrait nous permettre d'aller plus directement vers des éléments nous permettant d'identifier l'intuition individualiste dans les textes étudiés, et ce afin d'arriver à notre conclusion à moindre coût.

Nous emprunterons ici une méthode d'analyse de texte répondant à ces impératifs, et inspirée des méthodes d'analyse de texte assistée par ordinateur (Meunier *et al.*, 2005; Fayyad *et al.*, 1996), elles-mêmes conçues pour traiter des corpus textuels trop larges pour être lus. De la même façon que ces dernières nous donnent des outils permettant d'isoler les associations entre des sous-ensembles d'un corpus et certaines expressions lexicales ou syntaxiques, notre méthode doit nous permettre d'associer la présence de frontières avec la présence de l'intuition individualiste. Celle-ci procède selon des étapes successives. Premièrement, la sélection d'un cor-

pus de référence, qui contient les textes qui forment le contexte général des discussions sur le thème étudié, puis parmi celui-ci, d'un corpus d'étude, qui contient les textes sur lesquels l'hypothèse sera testée. On traduit ensuite l'hypothèse générale sous forme d'une hypothèse plus formelle, qui porte sur des éléments que l'on peut tester, tout en restant assez générique pour s'appliquer à tous les textes étudiés ; puis on conçoit le test lui-même. Enfin, on l'applique et on interprète les résultats.

Dans un premier temps, il s'agit de reconnaître les textes pertinents à notre étude. Ici, notre corpus de référence consiste en premier lieu en un ensemble de textes de tradition philosophique écrits dans le cadre des débats sur l'esprit étendu qui ont surtout émergé des réactions à l'article de Clark et Chalmers (1998), ou qui ont été reconnus *a posteriori* comme partie prenante dans ces débats (Hurley, 1998; Haugeland, 1993). Ces auteur·es se citent abondamment les un·es les autres, et se sont rencontrés souvent dans le cadre de conférences et de colloques, et empruntent les mêmes termes et les mêmes concepts pour parler d'esprit étendu. On retrouve dans ce groupe Andy Clark, Mike Wheeler, Mark Rowlands, Fred Adams, Ken Aizawa, Susan Hurley, Rob Rupert, etc. À cela, on peut ajouter des auteur·es qui se sont greffé au débat en empruntant le langage et les préoccupations des auteur·es de façon à cadrer explicitement leurs contributions dans les débats sur l'esprit étendu. Il s'agit souvent de collègues ou d'étudiant·es des premier·es (Michele Merrit, Shannon Spaulding, Zoe Drayson, Mirko Farina, ...), ou alors de chercheur·ses ayant une autre spécialité pertinente qui viennent apporter leur contribution au débat (Edwin Hutchins, Matt Barker, Amanda Barnier, ...). L'ensemble de ces contributions forme un contexte de discussion qui est aisément reconnaissable tant par le langage qui y est utilisé que par les questions qui y sont abordées.

Le corpus de référence a à la fois pour fonction de délimiter l'espace de discours sur lequel porte notre hypothèse et d'être le domaine à partir duquel on sélectionne les textes qui constituent le corpus d'étude. Ce dernier doit être constitué des textes parmi le corpus de référence qui sont réellement pertinents pour déterminer la véracité de l'hypothèse – en l'occurrence, celle de l'hypothèse H2. Il y a donc

un travail à faire pour bâtir le corpus d'étude. Dans ce cas en particulier, nous verrons dans le chapitre II que le premier tri nous donne trop de textes qui peuvent *a priori* paraître pertinents, mais qui ne le sont pas vraiment. L'essentiel de ce chapitre visera donc à justifier l'exclusion de certains textes, à savoir ceux qui ne sont ni pertinents pour établir H2, mais qui sont suffisamment importants pour qu'on ait à les traiter. Dans le chapitre III, nous tenterons d'établir H2 à travers l'étude de plusieurs critères qui sont particulièrement importants dans les débats sur l'extension de l'esprit. Enfin, la conclusion rassemblera des critiques à l'intuition individualiste qui émergent des chapitres précédents, et établira ainsi H3 – de sorte que, H2 ayant été démontrée, on aura aussi établi H1.

1.5 Théorie de la caractérisation

Avant de commencer cette investigation, il convient de faire quelques clarifications qui seront utiles dans la suite. La première clarification concerne le concept de « caractérisation », qui a pour objet les tentatives de donner une description de l'esprit se traduisant aussi par une extension définie. Cette clarification, comme on le verra au chapitre II, est particulièrement pertinente pour la construction du corpus d'étude. Cependant, elle précise aussi le sens de la question générale dans laquelle se situe ce mémoire – à savoir la question de l'extension, comprise comme entreprise de caractérisation.

On l'introduit premièrement pour se distancer de la notion de « critère », qui est utilisée par la grande majorité des auteur·es, nous semble trop limitée pour les besoins de notre recherche. Un critère s'exprime généralement par un ensemble de conditions nécessaires et suffisantes, lesquelles sont liées par des opérateurs logiques – c'est en tout cas la conception de « critère » la plus populaire dans la littérature autour de l'esprit étendu. Or, si on considère le critère comme une heuristique pour déterminer si quelque chose fait partie de l'extension de l'esprit, on constate qu'il constitue en fait un type particulier d'heuristique⁷, et qu'il y a

7. Formellement, il peut sembler que toute heuristique H peut être transformée en un critère

donc des heuristiques qui ne sont pas représentables sous forme de critère. Par exemple, il semble qu'on ne puisse dire de la caractérisation du cognitif par *cluster concept* de Von Eckardt qu'elle soit un critère du cognitif.

Que la structure du critère semble d'emblée exclure certains types de caractérisations est un problème en soi, mais on pourrait encore espérer qu'un critère qui exprimerait de façon succincte et formelle un concept beaucoup plus complexe – à la manière de l'*explicatum* de Carnap (1950), par exemple – puisse être une approximation suffisante pour contribuer à faire avancer la compréhension du sujet. Cependant, si la question est de savoir si la réalisation spatiale du concept en question possède des frontières, on peut penser que ce caractère formel puisse piper les dés. Le critère produit une catégorisation bivalente : soit quelque chose est cognitive, soit elle ne l'est pas. On peut donc s'attendre à ce qu'il en aille de même pour l'extension spatiale de cette catégorisation : le passage d'une aire où les entités contenues sont cognitives à une où elles ne le sont pas se fait sans transition, à travers une frontière.

Certes, les conditions nécessaires et suffisantes pourraient porter sur des propriétés de haut niveau qui se réalisent sur un espace flou, donc il y a moyen de produire une extension floue tout en employant la forme du critère, mais il n'en reste pas moins que celle-ci limite les moyens d'y arriver. De plus, non seulement cela suppose tout de même que ces propriétés de haut niveau ne soient pas définies de cette façon. Par ailleurs, l'idée que la forme du critère puisse moduler les attentes des chercheurs reste préoccupante : on peut s'attendre à ce qu'ils recherchent une réponse qui se conforme à ce que semble supposer le critère, c'est-à-dire une certaine bivalence.

Deuxièmement, on cherche quelque chose qui soit justifiable, et donc communicable. Il est commun que l'on emploie des concepts dont on ne saurait expliquer comment ils nous permettent de faire la discrimination qui nous permet de re-

de la façon suivante : on stipule une propriété P qui correspond au succès de l'heuristique H . Le critère est alors compris de la propriété P , qui fait office de condition nécessaire et suffisante. Cependant, dans un tel cas, ce n'est pas le critère qui fait le travail de discrimination, mais plutôt l'heuristique H – le critère est superflu.

connaître leur objet. Faute de savoir comment ils opèrent, on les dit souvent « intuitifs ». Cependant, qualifier un concept d'« intuitif » ne nous épargne pas le devoir de l'expliquer ou de le justifier lorsque les circonstances le demandent. Ainsi, dire du concept de « cognitif » qu'il s'agit d'un concept intuitif ne nous apporte pas d'information additionnelle à son effet – le besoin de l'élucider reste entier.

En ce sens, la caractérisation se distingue du concept, qui est aussi libéral dans la forme qu'y peut prendre l'information – puisque, là aussi, toute heuristique raisonnable peut fonctionner – mais qui n'est pas soumis à la contrainte de communicabilité, puisqu'il n'a qu'à se manifester dans le comportement de l'organisme qui l'emploie.

Troisièmement, il y a un seuil d'informativité que doit atteindre la caractérisation, en-dessous duquel on ne peut dire avoir capturé l'objet en aucun sens du terme. Là encore, le fait que la caractérisation doive servir à la question de l'extension nous permet de se donner une norme minimale en la matière : s'il n'y a pas suffisamment d'information pour déterminer, même de façon très schématique et très approximative, une extension de l'esprit, alors on n'a pas de caractérisation de l'esprit. Par ailleurs, la caractérisation ne peut être partielle – elle doit nous informer sur l'ensemble de son objet, sans quoi on n'a pas capturé le cognitif, mais seulement un phénomène à l'intérieur du cognitif.

Cependant, ce seuil minimal reste bas. L'échec des différentes tentatives de caractérisation suggère en effet que l'on doit rester modeste dans notre façon d'en développer de nouvelles. Aussi est-il improductif d'exiger d'emblée une caractérisation compréhensive, qui cernerait entièrement ce qui est spécifique au cognitif ou à l'esprit. Il est plus raisonnable d'exiger, comme condition minimale, une esquisse de certaines propriétés du cognitif qu'un schéma exhaustif

Ces trois contraintes – libéralité de la forme, justifiabilité et informativité – sont les contraintes fondamentales que notre question pose à la caractérisation, et elles sont aussi ce qui fait qu'on ne peut recycler sans les modifier les recettes conceptuelles

communes – critère, définition, concept ou explication carnapienne. On peut ainsi définir la caractérisation selon la formule faite plus haut : heuristique ou théorie justifiable qui nous permet de discriminer ce qui appartient à l’objet de ce qui ne lui appartient pas. De ce qu’on vient de dire, on peut tirer quatre conséquences.

Premièrement, de la contrainte d’informativité, on a tiré la conclusion qu’il nous fallait une caractérisation *extensionnelle*, c’est-à-dire qui puisse déterminer une extension au concept caractérisé. Ce que demande une caractérisation extensionnelle, c’est qu’elle puisse nous informer sur l’extension du concept étudié. Bien entendu, elle peut exploiter des propriétés physiques occupant un espace déterminé – e.g. tout ce qui est cognitif se réalise dans des tissus biologiques vivants (Adams et Aizawa, 2008). On pourrait aussi avoir des critères au niveau fonctionnel ou au niveau des processus, pourvu qu’ils produisent des contraintes qui les rendent reconnaissables empiriquement : par exemple, on dira de quelque chose que c’est de la mémoire si ça a la capacité de recevoir et de conserver de l’information. Enfin, même pour une propriété conçue au niveau intentionnel, il est possible de concevoir des contraintes dans l’implémentation qui nous renseignent sur l’extension. Le meilleur exemple est sans doute le contenu non-dérivé, que Adams et Aizawa (2001) avancent comme marque du cognitif. Si Adams et Aizawa ne donnent pas vraiment les raisons qui leur fait croire que le contenu non-dérivé doit se trouver dans le cerveau, Menary (2007) et Rowlands (2009) ont développé des arguments pour montrer qu’il se manifeste dans certains processus sous certaines conditions, et en ont produit un argument pour l’esprit étendu.

Deuxièmement, la caractérisation, à l’instar de l’explication carnapienne, appelle deux justifications : elle doit justifier, d’une part, qu’elle caractérise un objet réel et pertinent, et d’autre part, que cet objet est bien celui que l’on cherche, à savoir l’esprit ou le cognitif. Dans la mesure où la caractérisation peut n’être que minimalement informative, il s’agit souvent dans le second cas de justifier qu’un schéma très grossier de l’objet caractérisé puisse passer pour ce dernier, au moins dans certains contextes. On n’attend pas de la caractérisation qu’elle soit équivalente au concept intuitif caractérisé, et elle peut être une grossière simplification, du moment où les caractéristiques pertinentes à la question étudiée

sont préservées.

Troisièmement, que la caractérisation puisse être une simplification grossière ne veut pas dire que l'on doive se contenter de celle-ci. Cependant, une simplification peut servir de levier qui nous permette d'élucider davantage de caractéristiques du concept étudié, permettant du coup à la caractérisation d'être enrichie, ce qui peut à son tour permettre encore d'autres découvertes qui lui sont pertinentes. Il y a là une heuristique intéressante qui permet le développement du concept étudié au fur et à mesure que l'on enrichit sa compréhension.

Finalement, la caractérisation extensionnelle ne met aucune contrainte quant à la présence ou non de frontières du cognitif. L'extension peut bien sûr être limitée à un espace délimité par des frontières, mais tous les espaces ne sont pas délimitables par des frontières. Personne ne peut trouver précisément les frontières de l'écosystème des ours noirs du parc de la Gatineau (Degré, 2011), mais je sais que je ne m'y trouve pas quand je suis chez moi à Montréal. L'extension peut prendre la forme d'un sous-espace flou, ou alors subir une certaine dynamique, et changer sans cesse selon la situation, comme cela semble être le cas chez Clark (2008).

Bien que certains philosophes, les internalistes en particulier (e.g. Rupert, 2009; Adams et Aizawa, 2008), semblent supposer implicitement que l'extension de l'esprit doit avoir des bornes déterminées, rien ne semble empêcher des approches plus libérales. De fait, si on choisit une forme de caractérisation libérale comme un *cluster concept* ou une heuristique, on pourrait s'attendre à ce que le flou ou le dynamisme de la forme se transpose dans l'extension. L'hypothèse d'une frontière déterminée ne vient donc pas de soi, et le fardeau de montrer que le concept d'esprit en demande une revient à celle ou celui qui la suppose.

1.6 Usages du concept d'esprit – qui en a besoin ?

La deuxième mise au point concerne l'horizon d'application du concept d'esprit ou du cognitif.

Parce que l'essor de la question de l'extension est lié aux traditions des sciences cognitives et de philosophie de l'esprit, les discussions autour de l'esprit étendu ont surtout été faites dans l'optique de ces deux disciplines. C'est-à-dire que, hormis les questions plus métaphysiques, les préoccupations qui guident son investigation sont des préoccupations de sciences cognitives. Ainsi on demandera : Quel concept d'esprit correspond le mieux au concept intuitif des chercheurs des sciences cognitives ? (Von Eckardt, 1995) Quelle extension rend le mieux compte des succès scientifiques passés ? (Rupert, 2009) Quel critère a le plus de chances de cerner l'objet naturel ou réel étudié par les sciences cognitives ? (Adams et Aizawa, 2008) Quel concept est à même d'amener les sciences cognitives sur des avenues fructueuses ? Lequel risque-t-il au contraire de nous faire passer à côté d'une découverte ? (Clark, 2008)

Cette emphase sur les questions de sciences cognitives se traduit en un biais qui s'exprime selon deux dimensions. D'une part, certaines disciplines sont complètement négligées. L'économie, les sciences politiques, le droit, les études féministes, la sociologie, l'anthropologie, etc. sont presque complètement absents des débats sur l'extension de l'esprit. D'autre part, certaines disciplines sont abordées par la bande, mais dans un esprit un peu « colonialiste » : on y traite de ce que les thèses sur l'extension de l'esprit peuvent y apporter, mais pas du contraire. Ainsi, on parle parfois d'éthique (e.g. Clark et Chalmers, 1998; Drayson et Clark, à paraître; Anderson, 2008; Buller, à paraître), mais c'est pour juger des répercussions de l'esprit étendu, et jamais pour évaluer les contraintes que celle-ci peut apporter à notre conception de l'esprit et du cognitif.

Dans les deux cas, un biais internaliste qui a façonné l'histoire des disciplines des sciences humaines explique peut-être cette discrimination. Il semble en effet que les sciences humaines suivent le plus souvent, dans les dichotomies institutionnelles,

une séparation au niveau de l'unité qu'est la personne. Les sciences cognitives sont plus intéressées aux processus ou aux états mentaux qui composent le système cognitif ou la personne – elles s'intéressent au niveau sous-personnel – alors que les sciences sociales partent d'une unité première et insécable comme la personne ou l'agent·e rationnel·le et étudient leurs interactions. Les sciences humaines semblent donc avoir hérité d'un partage entre ce qui est interne à l'esprit et ce qui lui est externe, et les théoriciens de l'esprit étendu, formés dans cette tradition, la font refléter dans leurs préoccupations.

La situation dans laquelle est le débat actuel sur l'extension de l'esprit impose qu'on prenne en compte les sciences sociales, et ce pour deux raisons. D'une part, lorsque l'internalisme est remis en question, l'esprit en vient à s'étendre jusqu'à la sphère de la science sociale, et pénètre ainsi son domaine d'étude. Les sciences sociales ne font alors pas qu'utiliser un concept des sciences cognitives – elles étudient des phénomènes internes à l'esprit, selon le nouveau paradigme. Ainsi, chez Gallagher (2012), l'esprit prend une extension maximale, jusqu'à englober tout l'environnement social de l'agent·e. Il faut donc que la sociologie en entier devienne une étude du cognitif. C'est d'ailleurs un souci semblable qui amène Sutton (2004) à faire un cadre conceptuel pour l'étude de la mémoire où il inclut l'étude des sciences sociales.

Par ailleurs, quand bien même les sciences sociales n'auraient rien à dire sur le fonctionnement interne de l'esprit ou du cognitif, le débat de l'extension de l'esprit a depuis longtemps délaissé l'adéquation du concept à son objet au profit de spéculations sur son usage scientifique – c'est-à-dire que l'enjeu du débat n'est plus tant la substance désignée sous le concept d'esprit ou de cognitif que l'usage que l'on fait du concept. Barker (2010) montre bien comment les deux alternatives théoriques principales, celles de Clark (2008) et de Rupert (2009), sont indistinguables autant au niveau empirique que sur d'autres facteurs mesurables, de sorte qu'il est impossible de trouver une base solide pour appuyer l'une ou l'autre des deux positions. C'est pourquoi les échanges entre ces deux philosophes mettent plutôt l'accent sur des arguments sur l'influence qu'un concept ou un autre peut avoir sur le développement de la recherche.

Alors que le débat s'aventure sur des spéculations sur ce qui est encore à découvrir, on peut s'étonner que les sciences sociales n'aient pas encore été consultées. Ces disciplines reposent sur des concepts qui sont en grande partie tributaires de conceptions de l'esprit et du cognitif, et elles en ont extrait certains succès empiriques. Si l'on propose des changements radicaux à notre conception de l'esprit, leurs bases conceptuelles devront refléter ces changements, tout en s'accordant à ces succès empiriques. Or il est possible que certains changements aient des répercussions qui ne permettent pas de remplir cette dernière condition, et alors, il faudra décider si l'on doit garder notre conception révisée de l'esprit ou retourner à l'ancienne afin d'en conserver les succès empiriques. Dans tous les cas, les sciences sociales semblent devoir exercer une certaine contrainte sur l'élaboration du concept d'esprit, et il serait temps que les théoriciens de l'extension de l'esprit le reconnaissent.

Enfin, nous reviendrons au chapitre III sur le statut des intuitions dans le débat, mais il est pertinent de mentionner ici que les sciences sociales permettent d'aborder nos intuitions concernant l'esprit dans un angle différent. Par exemple, on a certainement des intuitions sur le type d'actes qui tombent sous notre responsabilité, qui ne sont sans doute pas les mêmes que nos intuitions concernant ce qui est cognitif – à résoudre la divergence, on peut clarifier notre intuition sur le cognitif. Ce nouvel angle a certainement le potentiel d'enrichir le débat.

1.7 Conclusion

Nous avons montré dans ce chapitre que la question de l'extension de l'esprit participait probablement d'une autre question, la question de la caractérisation de l'esprit, et nous avons choisi d'interpréter la précédente à la lumière de la première. Nous avons ensuite énoncé la problématique du mémoire à proprement parler : l'extension que prend le cognitif doit-elle obligatoirement se décrire à l'aide de frontières? Nous avons enchaîné avec la démarche, puis nous avons décrit succinctement la méthode d'analyse de texte qui sera utilisée dans la suite. Puis,

nous avons fait deux mises au point afin de mieux situer notre problématique. Premièrement, nous avons clarifié la notion de caractérisation, qui intervient dans la formulation de la question de l'extension. Deuxièmement, nous avons fait remarquer qu'il y avait eu une erreur dans la façon de traiter la question en ceci que les sciences sociales avaient été oubliées, et nous avons suggéré que cet oubli était dû à un relent d'internalisme qui se manifeste même chez les partisans de l'esprit étendu.

Ces clarifications ont une importance particulière à notre entreprise, en ceci qu'elles soulignent des biais qui affectent les débats sur l'esprit étendu. Or, dans la mesure où notre but sera de démontrer le biais en faveur d'une extension individualiste de l'esprit déterminée par des frontières, il importe d'être sensible à ces questions. En montrant le biais individualiste qui semble affecter le concept de critère et le choix des disciplines sensées informer les sciences cognitives, on explique d'une part en partie ce qui a mené la littérature à la position qu'on entend remettre en question, et on se donne l'occasion de corriger le tir et de partir sur une base plus neutre.

Dans la suite, nous tenterons de montrer par une revue de littérature comment ce biais individualiste se manifeste dans les critères du cognitif. Celle-ci commencera dans le chapitre suivant par les critères du cognitifs qui renvoient en fait à d'autres critères.

CHAPITRE II

REPOUSSER LA QUESTION

Ce chapitre porte essentiellement sur la construction d'un corpus d'étude – une exigence qui provient de la méthode, mais aussi de la nature même des questions que nous nous posons. Concernant la méthode, l'exigence vient à la fois du fait qu'il nous faut cibler les textes pertinents pour répondre adéquatement à nos questions et du fait qu'il nous faut nous restreindre dans la quantité de textes pour pouvoir tous les traiter. Concernant nos questions, d'une part, en effet, l'hypothèse H2 – qui veut que l'intuition individualiste soit toujours présente lorsque l'on propose une caractérisation du cognitif qui se présente dans un espace que l'on peut décrire adéquatement à l'aide d'une frontière – se limite à un certain contexte de discours, soit celui des débats actuels autour de l'esprit étendu. D'autre part, cette exigence pour l'hypothèse H2 vient du fait que l'hypothèse H1 – l'hypothèse générale de ce mémoire, selon laquelle on peut envisager de définir l'extension de l'esprit avec d'autres moyens que par des frontières – est elle-même contextualisée. En effet, on pourrait penser que H1 a une portée universelle, puisque rien dans l'énoncé ne fait référence à un contexte spécifique. Cependant, on ne poserait jamais H1 – ni d'ailleurs la question à laquelle H1 répond – s'il n'y avait à la base, dans les cercles qui étudient la question de l'extension du cognitif, une présomption particulière en faveur d'une extension définie par frontières. Aussi H1 et H2 portent-elles sur un contexte discursif particulier, duquel on doit extraire les textes du corpus d'étude, sur lesquels on fera notre étude, et qui correspond au corpus de référence.

Afin de remplir chacune de ces contraintes, il nous faut donc partir du corpus de référence, et y puiser les textes pertinents à l'évaluation de l'hypothèse H2. Aussi la méthode qui s'impose sera-t-elle une analyse de texte, prenant la forme d'une revue de littérature. En effet, il serait vain de chercher des auteur·es représentatif·ves des philosophes qui adhèrent ou n'adhèrent pas à l'intuition individualiste, puisque celle-ci est rarement explicite. Cependant, même si c'était le cas, notre hypothèse vaut pour le contexte actuel des débats, lequel n'est pas déterminé dans notre hypothèse. Aussi, pour affirmer qu'un·e auteur·e est représentatif·ve des philosophes soutenant le même point de vue quant à l'intuition individualiste, il faudrait avoir une connaissance pertinente de ce contexte. Or, si c'était le cas, on n'aurait sans doute pas à démontrer notre hypothèse – elle irait de soi. Il nous faut donc étudier toute littérature pertinente (ou à peu près) pour avoir un jugement informé sur la question.

Cette sélection, bien qu'elle puisse sembler triviale, est un processus important et très complexe dans le cas qui nous intéresse. Important, car lorsque l'on enquête sur un corpus, le choix de celui-ci détermine les réponses qu'on va trouver. Il faut donc s'assurer que ce soit le bon si on veut produire des résultats valides. Et complexe car, comme nous l'apprend la suite (section 2.1), les critères pour former notre corpus d'étude ne peuvent porter que sur des caractéristiques qui ne sont pas toujours explicitées dans le texte.

2.1 Critères de sélection

Quels critères emploie-t-on pour construire le corpus d'étude? D'une part, la littérature doit porter sur l'esprit étendu dans sa forme actuelle – telle que défendue par Clark, Wheeler, Rowlands, etc. En effet, d'une part, c'est, à notre connaissance, le seul espace de discours dans la philosophie occidentale actuelle qui traite réellement de l'extension de l'esprit. Ensuite, on juge que les auteur·es ayant abordé le thème auparavant ne partaient pas du même contexte intellectuel – la question de l'extension ne se posait pas dans l'optique de produire une

extension déterminée, de sorte que le concept de frontières ne remplissait pas le même rôle. D'autre part, elle doit nous permettre de tester notre hypothèse – on doit donc pouvoir évaluer, minimalement, si l'intuition individualiste est présente et si elle mène à postuler des frontières du cognitif.

La première condition consiste à rester à l'intérieur du corpus de référence : y appartiennent les textes qui se positionnent explicitement dans ce débat, qui le font implicitement en citant les sources classiques de l'esprit étendu (Clark et Chalmers, 1998; Adams et Aizawa, 2008), ou qui sont considérés comme faisant partie intégrante du débat par la communauté des chercheurs qui étudient la question de l'extension (l'exemple par excellence est celui de Haugeland, 1993, qui a été intégré dans le canon *a posteriori*). Par les citations et par le vocabulaire employé, il est généralement très facile de déterminer si un texte appartient ou non au corpus de référence.

Par ailleurs, nous prendrons ici un raccourci consistant à ne prendre qu'un échantillon de l'œuvre des auteur·es du corpus de référence. En effet, on peut généralement supposer qu'il y a une certaine continuité dans la pensée d'un·e auteur·e, qui fait que sa position par rapport à un à un moment est représentative de sa position dans le reste de son œuvre.

La seconde condition pose davantage de difficultés : comment, en effet, peut-on déterminer si un·e auteur·e est susceptible de souscrire à l'intuition individualiste ou à un domaine du cognitif ceint par des frontières avant d'avoir déterminé s'il souscrit à ces idées ou à leur contraire? La chose est impossible, mais on peut en revanche restreindre l'ensemble de la littérature pertinente en identifiant des conditions sous lesquelles un discours ne saurait nous permettre de tester notre hypothèse.

On peut ainsi proposer des conditions qui circonscrivent des caractéristiques essentielles à notre analyse. Ainsi, on a dit qu'on devait évaluer notre hypothèse, laquelle portait sur la relation entre la présence de frontières du cognitif (dans le sens qu'on a défini) et la souscription à l'intuition individualiste. Dès lors, on peut

proposer des conditions pertinentes à l'évaluation d'un texte par rapport à l'une et à l'autre.

Ainsi, en premier lieu, on a besoin d'un texte ou d'un corpus qui propose une caractérisation de l'esprit. Ce besoin vient du fait que l'on a besoin d'une extension du cognitif – en effet, impossible sans cela de dire si on a des frontières, et donc impossible de vérifier notre hypothèse. Celle-ci ne saurait être arbitraire, sans quoi elle serait sans intérêt : il doit donc y avoir une caractérisation du cognitif qui justifie l'extension proposée. Ceci dit, en pratique, il est souvent plus facile de dire si un texte propose une caractérisation qu'une extension. En effet, l'auteur·e, ayant ses propres objectifs, ne s'attarde pas toujours suffisamment sur l'extension pour qu'on puisse clairement trancher sur la présence de frontières. On peut cependant parvenir à une réponse en appliquant les principes de la caractérisation. Autrement dit, certaines propositions quant à l'extension se manifestent en fait en propositions de caractérisations, de sorte que pour avoir des informations quant à l'extension, un certain travail d'assemblage est requis. Bref, on ne saurait que faire d'une extension sans caractérisation, mais une caractérisation peut toujours cacher une extension. Le critère pertinent ici, pour ne pas exclure indûment des textes pertinents, est donc la présence de caractérisation.

En second lieu, il nous faut une caractérisation qui soit suffisamment riche pour nous permettre de déterminer si l'intuition individualiste est sous-jacente dans le texte ou l'ensemble de textes étudié. Or une caractérisation doit toujours mener vers une extension : si on ne peut en déterminer une, c'est qu'il nous manque la théorie pour y parvenir. Par exemple, elle reposera sur une notion qui n'est volontairement pas définie. On dira alors qu'elle est *incomplète*. Ce type de caractérisation peut parfois être utile lorsqu'un·e auteur·e veut faire avancer son argumentation sans pour autant produire une théorie complète du cognitif. Par exemple, disons qu'un philosophe veut s'opposer à une certaine théorie de l'extension du cognitif. Elle produit une caractérisation incomplète, qui fait rapporter le cognitif à une notion dont elle laisse l'explication en suspens. Si elle possède une intuition quant à la localisation et l'extension de cette notion, elle peut justifier des conjectures sur la localisation et l'extension du cognitif en se basant sur cette

intuition.

Ce genre de stratégie est à distinguer d'une stratégie où les notions sont expliquées exhaustivement. Certes, il n'y a pas d'explication tout à fait complète, parce qu'en suivant le fil des définitions, on finit toujours par tomber sur des concepts qui ne sont pas définis. Cependant, si ces concepts réfèrent aux lieux communs ou *topoi* de la discipline pertinente (ici, la philosophie de l'esprit ou des sciences cognitives), la personne qui lit en a déjà une certaine compréhension (Aristote, 2007, 1358a) – il n'est donc pas nécessaire de les définir pour produire la compréhension escomptée. Par exemple, la première stratégie serait analogue à définir la justice comme « ce que tous entendent par justice » et la seconde serait analogue à définir la justice comme étant un état où tous les échanges sont équitables en valeur, et à justifier cette définition par le fait que tous entendent la justice ainsi.

Ainsi, lorsqu'il s'agit de déterminer quelles intuitions sous-tendent une conception du cognitif, les critères incomplets rendent la tâche plus difficile. Si l'auteure a telle intuition concernant une notion, serait-ce parce qu'elle a en fait certaines idées quant à la théorie qui pourrait l'expliquer ? Si elle emploie telle intuition, serait-ce parce qu'en quelque part, elle s'attend à ce qu'un certain programme de recherche finisse par fournir cette explication, et que ce programme de recherche est basé sur cette intuition ? Les conjectures qu'on fait lorsque la théorie manque ne sont pas toujours anodines, et l'incomplétude d'un critère tend à obscurcir la voie. Au final, on ne saurait arriver à une conclusion concernant notre hypothèse – l'association entre la présence de frontière et l'intuition individualiste – dans le cas des critères incomplets.

Pour y arriver, il faudrait les « compléter » en remplaçant les notions laissées en suspens par des notions explicitées. Cependant, on verra dans le prochain chapitre que le matériel théorique nécessaire pour cette opération se trouve généralement déjà dans les critères complets. Autant donc, pourrait-on dire, se débarrasser des critères incomplets pour se concentrer sur les critères complets. Néanmoins, les critères les plus notoires et les plus discutés sont souvent des critères incomplets, et leur incomplétude n'est pas toujours évidente. Aussi convient-il de commencer

l'étude des critères du cognitif par ceux-ci, afin de mettre la table pour l'étude des critères complets.

Dans les prochaines sections, nous nous pencherons donc sur le principe de parité (Clark et Chalmers, 1998; Clark, 2008; Wheeler, à paraître; Rowlands, 2010b; Gallagher, 2011), la manipulabilité mutuelle (Kaplan, 2012), l'accord avec l'évidence empirique (Clark, 2008; Rupert, 2009) et le contenu non-dérivé (Adams et Aizawa, 2008). Tous sont traités comme des critères du cognitif dans au moins une partie de la littérature. Notre travail sera de montrer qu'ils ne sont pas des critères complets – et donc pas des caractérisations du cognitif, au sens où on l'a défini.

Il est peut-être pertinent de mentionner finalement que les thèses qui seront abordées ne sont pas nécessairement présentées comme des critères du cognitif ou de l'extension du cognitif par leurs auteur-es. Cependant, elles sont toujours, à un moment ou un autre, interprétées et discutées comme tel dans la littérature. Comme notre enquête porte sur les critères du cognitif, ce sont ces interprétations, et celles-là seulement qui nous intéressent. Aussi, dans ce chapitre et dans celui qui suit, l'angle sous lequel on aborde les différentes thèses de différents auteur-es n'est-il pas toujours le plus fidèle à l'énonciation originale, et ce parce que l'angle choisi est celui qui correspond le mieux à notre tâche, soit la recherche et à l'évaluation des critères du cognitif.

2.2 Le principe de parité

Le premier en lice est bien entendu le principe de parité. Sa première formulation nous a été donnée par Clark et Chalmers (1998, 8) :

If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process.

Ce principe s'illustre à travers une série d'exemples et d'expériences de pensée, dont la plus célèbre va comme suit. Inga, une personne sans handicap psychologique apparent, veut se rendre au Musée d'art moderne (MoMA) de New York. Elle se rappelle avoir noté mentalement que le musée en question se situe sur la 53^e rue, et se met donc en route dans cette direction. Otto, qui est frappé de la maladie d'Alzheimer, n'a pas les mêmes ressources qu'Inga : il arrive très difficilement à mémoriser de nouvelles choses. Pour pallier à ce problème, il a toujours sur lui un petit carnet dans lequel il note méticuleusement tout ce qu'il doit retenir. Lorsque l'envie lui prend d'aller au musée, Otto sort son calepin, voit qu'il se situe sur la 53^e rue, et prend la direction du musée.

Dans le contexte de ce problème, le carnet joue pour Otto le même rôle que jouent pour Inga les circuits neuronaux qui réalisent sa mémoire. L'un et l'autre sont donc fonctionnellement équivalents, c'est-à-dire que leur similitude se trouve dans leur profil fonctionnel : si l'organisme possède les interfaces nécessaires pour interagir correctement avec l'un et l'autre, ils sont capables de remplir la même fonction. Dans cette mesure, si on dit d'une des deux mémoires qu'elle réalise un processus cognitif, on devrait par parité pouvoir le dire de l'autre.

2.2.1 Quatre versions du principe de parité

Le principe de parité repose à la fois sur un fond de fonctionnalisme¹, qui se décline en plusieurs variantes, et sur cette intuition qui nous amène à croire *grosso modo* que les traitements ou échanges d'information se produisant dans le cerveau forment des processus cognitifs. Cette intuition – l'intuition *internaliste* – peut être interprétée de diverses manières. Son rôle est de nous donner une base de processus qui sont plus ou moins indubitablement cognitifs. Ainsi, notre intuition du cas d'Inga doit nous accorder qu'il est plus ou moins indubitable que le rappel de mémoire est une opération cognitive. Le rôle du fonctionnalisme est de donner

1. Par fonctionnalisme, on entend ici une thèse sur la nature ou la définition des phénomènes, et notamment des phénomènes cognitifs ou mentaux : une condition nécessaire et/ou suffisante pour être x est de posséder le patron fonctionnel y .

une base de comparaison, ou une dimension ontologique, grâce à laquelle on peut établir l'équivalence entre différentes opérations ou processus. En l'occurrence, il s'agit de la dimension fonctionnelle : si deux choses sont équivalentes en cette dimension, elles sont équivalentes quant à leur statut cognitif. En associant les deux thèses, on obtient ceci : si un processus y a le même patron fonctionnel qu'un processus x dont on sait qu'il est cognitif (en vertu de l'intuition internaliste), alors y est cognitif.

Le choix d'une variante affecte évidemment la portée du principe de parité. Le fonctionnalisme intervient ainsi dans la définition du cognitif : un fonctionnalisme fort dira que les processus cognitifs se définissent essentiellement en terme de leur patron fonctionnel, alors qu'un fonctionnalisme plus faible dira qu'ils se définissent en partie par leur patron fonctionnel, ou alors que seulement certains processus cognitifs se définissent comme tels. Si on a un fonctionnalisme fort, alors l'usage du principe de parité permet à lui seul de déterminer qu'un processus est cognitif (Chalmers, 2008, xv). La confiance qu'on a que y est cognitif est alors équivalente à la confiance qu'on place dans notre intuition que x l'est. Si on a un fonctionnalisme faible, le patron fonctionnel est pertinent à déterminer le statut cognitif d'un processus, mais n'est pas nécessairement seul à l'être. Il est donc possible que, bien que x et y aient des patrons fonctionnels identiques, l'un soit cognitif et l'autre pas – à condition qu'il y ait une différence pertinente. Dans ce cas, il revient donc à la personne qui défend une conception internaliste (*brainbound*) du cognitif de démontrer qu'il existe une telle différence qui soit de nature à exclure les processus externes du cognitif. Autrement dit, le fardeau de la preuve lui revient.

Par ailleurs, bien que la formulation originale du principe de parité repose sur l'intuition internaliste, on pourrait envisager la remplacer par un critère ou une heuristique de découverte – c'est du moins ce que soutient Wheeler (à paraître). Cependant, on peut se demander si, sans intuition internaliste, on a toujours bien un principe de parité. Si on a un critère indépendant de la localisation, alors il s'applique de la même façon à l'intérieur comme à l'extérieur de la boîte crânienne – inutile, donc, d'essayer d'établir des correspondances de patron fonctionnel entre processus internes et externes, puisque l'on sait déjà qu'ils ont la même barre à

franchir pour être considérés comme cognitifs. Si on a une procédure indépendante pour mesurer une distance avec la vitesse de la lumière, il est superflu (et moins précis) de la déterminer en utilisant l'étalon du Bureau international des poids et mesures. Il en va de même avec le caractère cognitif d'un mécanisme : un critère doit rendre l'étalon obsolète. Wheeler semble reconnaître ce problème, mais offre une réponse sibylline :

The right response here (I think) is to point out that the strategy under consideration may be explicated using a perfectly reasonable notion of parity, just so long as parity is conceived not as parity with the inner simpliciter, but rather as parity with the inner with respect to a locationally uncommitted account of the cognitive. (Wheeler, à paraître, chp. 3, 14)

On voit mal en quoi cette réponse peut satisfaire celui ou celle qui craint que le principe de parité ne tombe dans la superfluité. Cependant, on peut trouver d'autres justifications. Peut-être Wheeler pense-t-il que le principe de parité ne fonctionne pas comme une heuristique, mais plutôt comme une contrainte sur le critère ou l'heuristique qui doit remplacer l'intuition internaliste : celui-ci devrait respecter le fait que, pour deux mécanismes fonctionnellement équivalents, le statut cognitif doit être le même. Par ailleurs, même si Wheeler ne semble pas considérer cette possibilité, il peut être plus facile de déterminer le statut cognitif de certains mécanismes que d'autres. Dans les cas où l'usage du critère ou de l'heuristique est plus difficile, on peut peut-être utiliser le principe de parité comme heuristique palliative et ainsi connaître indirectement le statut cognitif du processus. Bref, on ne peut exclure ici que le principe de parité puisse se baser sur un critère ou une heuristique informée théoriquement, comme le suggère Wheeler, plutôt qu'avec l'intuition internaliste.

En somme, on peut, de ces deux variations, déduire qu'il y a 4 variétés du principe de parité. Clark et Chalmers défendent une version faiblement fonctionnaliste avec l'intuition internaliste. Cependant, on doit noter que pour les deux auteurs, le principe de parité doit être considéré comme une pompe à intuition (Clark, 2008) : il doit montrer que le consensus autour de l'internalisme ne repose pas sur des bases solides. Ensuite figure ce qu'on pourrait appeler la version « prototypique »

du principe de parité, que l'on retrouve chez Gallagher et Crisafi (2009); Gallagher (2011, 2012) et, à titre d'homme de paille, chez Rupert (2009). Celle-ci est fortement fonctionnaliste et emploie aussi l'intuition internaliste. Wheeler, de son côté, emploie une version fortement fonctionnaliste basée sur un critère informé théoriquement. Enfin, on chercherait en vain un philosophe qui défend une version faiblement fonctionnaliste basée sur un critère informé théoriquement, mais c'est peut-être parce qu'un fonctionnalisme faible est assez peu contraignant sur un critère informé théoriquement, et qu'en ce sens, il est plus à même de devenir redondant.

2.2.2 L'intuition internaliste et les Martiens

Dans le cas de Wheeler, qu'on pense qu'il ait raison ou tort d'affirmer que son argumentaire repose sur une forme du principe de parité, il n'en reste pas moins que la variété qu'il épouse repose sur un critère ultérieur. Il l'admet d'ailleurs explicitement, laissant le soin de le développer au domaine de l'intelligence artificielle (Wheeler, à paraître, chp. 5). Dès lors, sa caractérisation du cognitif doit être considérée comme incomplète, et ne saurait faire partie d'un corpus approprié pour l'évaluation de notre hypothèse. Restent donc les deux variétés qui reposent sur l'intuition internaliste. Mais qu'est-ce que cette intuition internaliste ?

Une expérience de pensée fort élégante de Sprevak (2009) laisse croire que cette intuition n'a aucun contenu – au sens où elle a certes une extension, mais aucune intension – et que cette vacuité crée des situations ambiguës lorsque l'on applique le principe de parité pour déterminer l'extension de l'esprit. Si cette conclusion s'avérait, le seul moyen de sauver le principe de parité comme caractérisation du cognitif serait de clarifier l'intuition internaliste en employant une théorie de l'intuition internaliste. Le principe de parité serait donc incomplet.

Imaginons un Martien humanoïde qui, comme nous, possède un cerveau encasté dans une boîte crânienne. Comme pour nous, ce cerveau serait un centre de traite-

ment important qui dirige les mouvements du corps, le traitement et la production du langage, ainsi que la prise de décision. Bref, ce cerveau n'aurait peut-être pas le même patron fonctionnel que le nôtre – par exemple, il n'aurait pas les mêmes biais cognitifs – mais en gros, cet organe aurait un rôle à peu près équivalent au nôtre. Sprevak invoque un principe de non-anthropocentrisme (qu'il appelle « l'intuition martienne ») pour dire que l'on doit accorder le même traitement au cerveau martien qu'au cerveau humain – ce qui semble intuitivement correct. Et donc, on aura pour le cerveau martien la même intuition internaliste que pour le cerveau humain.

Cette intuition tient même si le martien utilise des mécanismes radicalement différents des nôtres. Par exemple, on peut imaginer que pour inscrire des informations en mémoire, ce cerveau utilise un système de marques d'encre et de lecture optique – une sorte d'analogue à l'écriture. Le partisan de l'esprit étendu est prêt à accepter cette idée, puisqu'il ou elle accepte aussi que le carnet d'Otto soit cognitif. Mais on peut pousser l'idée plus loin. Mon ordinateur possède plein de programmes qui ne me sont pas particulièrement utiles – par exemple, j'ai un programme qui me permet de calculer les dates du calendrier maya. Je n'ai évidemment pas tendance à dire qu'il fait partie de mon système cognitif, même si je ne me sépare jamais de mon ordinateur – je ne me sers jamais de ce programme. Imaginons cependant que ce programme est intégré dans le cerveau de notre martien. Il n'est pas plus porté que moi à s'en servir – surtout sur Mars, où les révolutions des astres sont différentes – pourtant, en vertu de l'intuition internaliste, on voudra dire qu'il s'agit d'un processus cognitif. Et donc, en vertu du principe de parité, le programme sur mon ordinateur fait aussi partie de mon système cognitif.

Sprevak pense que le principe de parité est une conséquence du fonctionnalisme dans sa version forte, et utilise cette expérience de pensée comme *reductio ad absurdum* contre l'un et l'autre. L'expérience de pensée fait ressortir la fragilité de l'intuition internaliste : si on accepte l'intuition martienne, on est pris pour admettre que tout et n'importe quoi peut être considéré cognitif, du moment qu'on peut le placer dans une boîte crânienne.

Sprevak a beau présenter son argument comme une conséquence directe du fonctionnalisme et de l'intuition martienne, l'argument, pour fonctionner, a aussi besoin de l'intuition internaliste. Aussi ne porte-t-il pas contre quelqu'un qui, comme Wheeler, n'y souscrit pas, et ce même si ce dernier est fonctionnaliste.

Clark et Chalmers, de leur côté, utilisent certes le principe de parité, mais avec un fonctionnalisme faible, et comme une pompe à intuition. Sprevak prétend que son argument vaut pour tout fonctionnalisme qui conserve l'intuition martienne, mais le fonctionnalisme faible que Clark (2008) emploie pour le principe de parité ne la conserve que sous réserve. Il ne lui permet pas d'affirmer l'extension de l'esprit comme une conclusion du principe de parité, car ce qui rend un processus cognitif pourrait ne pas seulement être son patron fonctionnel. Ce qu'il fait, c'est plutôt nous montrer que nous n'avons pas de raisons sérieuses de limiter l'extension de l'esprit au cerveau. Ainsi, le même constat modal s'applique au cerveau du martien : nous n'avons aucune raison sérieuse d'affirmer qu'il n'est pas cognitif. Cependant, le fonctionnalisme faible ne permet pas d'affirmer l'équivalence des deux cerveaux ; aussi n'est-il pas raisonnable d'en faire le lieu de l'intuition internaliste.

Reste la version prototypique, qui repose à la fois sur le fonctionnalisme fort et l'intuition internaliste. Pourtant, même dans cette variété, le principe de parité ne tombe sous la *reductio* de Sprevak qu'à la seule condition d'adopter le même type d'intuition internaliste qui est adopté dans l'exemple du martien. Dans ce type d'intuition, ce qui importe est la localisation du processus : si c'est dans le cerveau, alors c'est cognitif – autrement dit, la définition de cognitif dépend de ce qui se trouve à l'intérieur du cerveau. Si on veut soutenir cette intuition sans être anthropo-chauviniste, on doit admettre que, semblablement, ce qui est dans le cerveau du martien est cognitif – d'où le problème que Sprevak souligne. Mais on peut aussi concevoir l'intuition internaliste comme l'observation d'une contingence : ce qui est dans le cerveau est cognitif non pas en vertu du fait d'être dans le cerveau, mais en vertu d'une autre raison qui nous échappe. Dans ce cas, lorsque l'on passe du cerveau humain au cerveau du martien, on n'est pas obligé d'accepter que tout ce qui s'y trouve est cognitif : si un mécanisme du cerveau

martien n'a pas d'équivalent fonctionnel dans le cerveau humain, il n'y a aucune raison de penser que l'intuition internaliste s'y applique.

Il existe des définitions de l'intuition sur lesquelles nous reviendrons au chapitre III, mais, *grosso modo*, une intuition en philosophie est généralement comprise comme étant une sorte de concept dont on peut reconnaître les instances et que l'on peut utiliser, mais que l'on ne peut pas énoncer. Dans ce cas, l'interprétation que fait Sprevak de l'intuition internaliste l'amène à la réduire à une règle sommaire, qui devient alors facilement énonçable :

Règle internaliste. Tout traitement d'information dans le cerveau est cognitif *en vertu* du fait d'être dans le cerveau.

Sprevak transforme ainsi l'intuition en critère suffisant : au moins une partie de ce qui est cognitif l'est *en vertu* d'être intracrânien.

Il n'est pas surprenant que ce critère aille à l'encontre du fonctionnalisme fort. En effet, le fonctionnalisme fort veut que le statut cognitif ne dépende que du patron fonctionnel. Or ici, le statut cognitif dépend de la localisation – du fait d'être dans le cerveau : les processus concernés ne reçoivent donc pas leur statut exclusivement en vertu de leur patron fonctionnel.

La *reductio* de Sprevak n'est donc pas fausse, mais ses termes sont inhabituels. On entend plus volontiers l'intuition individualiste comme une observation sur la localisation des processus cognitifs, et non pas comme un critère suffisant du cognitif. En l'occurrence, il est très peu charitable d'attribuer cette thèse à Gallagher, ou à tout autre partisan de la version prototypique du principe de parité.

En somme, la *reductio* de Sprevak n'est dévastatrice pour aucune des versions du principe de parité, parce qu'elle interprète mal l'intuition internaliste. Dès lors, l'intuition internaliste ne peut être vide au sens où Sprevak entendait le montrer.

2.2.3 Fonctionnalisme et thermostat humain

Cependant, si l'expérience du martien ne vient pas à bout de l'intuition internaliste, elle nous pointe vers un autre problème du fonctionnalisme : un système dûment cognitif peut accomplir des fonctions qui n'ont rien de cognitif. Ceci devient clair lorsque l'on modifie un peu l'expérience de pensée. Prenons Joséphine, qui est une experte des civilisations précolombiennes. Joséphine a appris par cœur le fonctionnement des calendriers mayas, et est capable d'en calculer les dates – comme mon ordinateur, et comme le martien de Sprevak. Elle ne s'est jamais servie de cette faculté qu'elle a apprise il y a très longtemps, mais pourrait facilement s'en servir n'importe quand. Pour mousser l'exemple, admettons que Joséphine a subi un ACV et qu'elle a oublié avoir appris cette compétence, mais qu'elle serait tout de même capable de la réaliser si elle ou quelqu'un d'autre la persuadait de tenter la chose. La compétence de Joséphine est donc à peu près fonctionnellement équivalente à un programme installé sur mon ordinateur, dont j'ignore l'existence jusqu'à ce que je consulte les manuels qui sont installés sur mon disque dur. On peut pousser l'intuition plus loin avec une tâche encore plus simple et plus bête. Prenons Jeannot, qui a la compétence de savoir, en sentant l'air d'une pièce, de savoir s'il faut partir ou fermer le calorifère pour garder une chaleur confortable – compétence fonctionnellement équivalente à celle d'un thermostat.

A priori, la fonction effectuée n'est pas cognitive, et donc le mécanisme qui la réalise ne devrait pas l'être non plus. Cependant, dans le contexte des expériences de pensée de Joséphine et de Jeannot, notre intuition va dans le sens inverse : même si les tâches en soit ne sont pas cognitives, comme Joséphine et Jeannot utilisent leur cognition pour les accomplir, alors le processus est cognitif. Si on accepte cette intuition, on reconnaît donc que le système cognitif peut accomplir des fonctions qui ne sont pas cognitives. Or on sait aussi que des fonctions qui ont généralement la réputation d'être cognitives peuvent très bien être répliquées dans des médiums qui ne sont pas des systèmes cognitifs (comme l'économie de la Bolivie, pour reprendre l'expérience de pensée de Block, 1978). Dès lors, non seulement a-t-on des intuitions qui suggèrent que le profil fonctionnel n'est pas suffisant (puisque'il pourrait se matérialiser dans à peu près n'importe quel médium), mais

il n'est pas non plus nécessaire (puisque un humain peut utiliser sa cognition pour émuler un thermostat). Tout cela semble nous amener à penser qu'il faille laisser tomber le fonctionnalisme (faible ou fort) pour la caractérisation du cognitif.

En réponse à Adams et Aizawa (2008) qui prétendent que le cognitif devrait se caractériser par des particularités fonctionnelles de bas niveau, comme celles affichées par les tissus neuronaux, Clark (2008) propose de faire la distinction entre fonctionnalisme de gros grain et de petit grain, et choisit de défendre la première version. Pour répondre à la critique que nous venons d'émettre tout en conservant le fonctionnalisme, on peut suggérer que le cognitif doive se décrire à un grain supérieur à celui de la simple tâche. On peut ainsi penser que le système est cognitif, entre autres choses, parce qu'il a la capacité d'apprendre des tâches et de les intégrer dans son économie cognitive – le niveau d'analyse pertinent est donc plus du niveau du système que du niveau de la tâche.

Cependant, définir le patron fonctionnel du système en entier est une tâche fort complexe, qui reviendrait en soi à une caractérisation du cognitif, rendant ainsi le principe de parité superflu. Il faut donc que ce soit un niveau intermédiaire, mais ni Clark, ni personne d'autre ayant employé le principe de parité ne donne d'information suffisamment précise pour le déterminer. Or sans cette information, non seulement est-il difficile d'appliquer le fonctionnalisme, mais le contenu pertinent de l'intuition internaliste est difficile à définir : on ne sait plus vraiment ce qu'on a intuitivement jugé cognitif en vertu du fait d'être interne, puisque l'on ne sait pas à quel niveau fonctionnel regarder. Le principe de parité semble donc effectivement incomplet, même si ce n'est pas exactement pour les raisons que donne Sprevak.

Ainsi, c'est peut-être pour éviter d'avoir à clarifier ce qu'il évoque quand il parle de fonctionnalisme de gros grain que Clark introduit sa version du principe de parité comme simple pompe à intuition plutôt que comme critère du cognitif. Gallagher semble être seul à prendre la pierre, ici – du moins, dans la mesure où il endosse entièrement l'argument du principe de parité. Mais même ici, il faut mettre un bémol, car Gallagher n'endosse pas le fonctionnalisme des Wheeler, Clark et cie :

I think the basic strategy here is to show that most of the objections to the extended mind hypothesis are too closely tied to the conception of the mind that the hypothesis is rejecting – or ought to be rejecting. That is, we need to conceive of the mind, not on a functionalist interpretation (*pace* Clark, Wheeler, and many others), but as enactively generated in the specific interactions of organism environment (where environment is social as well as physical). The mind, not as a collection of propositional attitudes, representations, mental states, etc., all of which have supposedly intrinsic or non-derived content (Adams et Aizawa, 2008), but the mind as an enactive and emotionally embedded engagement with the world through which we solve problems, control behavior, understand, judge, explain, and generally *do* certain kinds of things. (Gallagher, 2011, 8)

Si Gallagher n'endosse pas le fonctionnalisme, il n'endosse certainement pas le fonctionnalisme fort ; s'il n'endosse pas le fonctionnalisme fort, comment peut-il endosser la version prototypique du principe de parité ? De fait, Gallagher reconnaît l'association du principe de parité avec le fonctionnalisme, et semble préférer pour cette raison qu'on base la thèse de l'esprit étendu sur une conception éactive de l'esprit : dans un article sur l'intentionnalité éactive (Gallagher et Miyahara, 2012, §2.1), il emploie même les difficultés qu'engendre le principe de parité comme une preuve à l'encontre d'une conception néobéhavioriste de l'intentionnalité. Si Gallagher emploie le principe de parité pour avancer son argument pour une extension de l'esprit dans le tissu social (son « *socially extended mind* »), c'est donc sans doute parce qu'il pense que les partisans fonctionnalistes de l'esprit étendu seront les plus réfractaires à son argumentation.

Bref, personne ne semble soutenir une version du principe de parité basée sur l'intuition internaliste comme principe justifiant l'extension de l'esprit. Les partisans du principe de parité tombent donc sous deux catégories : soit on ne se base pas sur l'intuition internaliste (et alors, on peut se demander si on a bel et bien affaire au même principe), et on renvoie à un autre critère, soit on utilise le principe de parité comme une pompe à intuition, qui doit préparer la lectrice ou le lecteur pour un argument plus solide. Dans tous les cas, la caractérisation se trouve ailleurs.

2.3 Manipulabilité mutuelle

En plus des raisons mentionnées dans la section précédente, l’impopularité de la version forte du principe de parité s’explique en partie par le fait qu’elle répond assez mal au problème de la surextension (*cognitive bloat*) – un des problèmes de l’esprit étendu qui est le plus souvent mentionné (e.g. Adams et Aizawa, 2008; Allen-Hermanson, 2012; Clark et Chalmers, 1998). Si on adopte une thèse externaliste qui, comme le principe de parité, propose de partir de ce qu’on est le plus porté à dire cognitif – les mécanismes du cerveau – et à étendre sa démarcation en vertu d’une parité fonctionnelle ou d’un autre principe, on en vient souvent à ne plus savoir où cela doit arrêter. Ainsi, certains équipements sont susceptibles d’étendre mon esprit en remplissant des fonctions qui sont aujourd’hui effectuées par ma matière grise, mais pour différentes raisons, je n’irais pas dire qu’ils appartiennent à mon système cognitif. Même si j’avais un accès me permettant d’effectuer des calculs sur un superordinateur de la défense nationale américaine, j’aurais du mal à considérer cette machine comme une partie de mon système cognitif.

C’est un peu la motivation derrière les quatre critères additionnels que proposent Clark et Chalmers (1998) : (1) que la ressource soit une constante dans la vie de l’organisme – qu’elle fasse partie de son mode de vie ; (2) qu’elle soit accessible facilement ; (3) qu’on lui fasse automatiquement confiance (dans le cas d’une croyance, qu’elle soit automatiquement endossée) ; et (4) que cette confiance vienne de l’expérience passée (dans le cas d’une croyance, qu’elle ait été endossée dans le passé). Cependant, ces critères sont un peu arbitraires – ils ne fonctionnent pas trop mal pour l’exemple de Otto et Inga, mais ne semblent pas particulièrement appropriés pour d’autres exemples. Ainsi, il arrive qu’on ait à réfléchir longuement pour se remémorer des événements qui se sont passés il y a très longtemps – pour ce faire, on peut peut-être tester diverses associations d’idées dans l’espoir de faire resurgir des traces de celui-ci, et tenter d’en faire ressortir un récit cohérent. L’accès qu’on a à cette mémoire n’est certainement pas constant, et le travail impliqué pour la réactiver est trop grand pour que l’on puisse qualifier l’opération de « facile ». Cela viole donc les deux premiers critères, mais cela ne nous empêche pas d’avoir

la ferme intuition qu'il s'agit d'un travail cognitif. Peut-être est-ce d'ailleurs une des raisons qui explique que Clark et Chalmers n'aient pas développé sur cette voie.

Dans l'optique d'offrir une réponse externaliste au problème de la surextension, Kaplan (2012) propose d'utiliser le critère de manipulabilité mutuelle, tel qu'il est utilisé par Craver (2007a,b) pour déterminer les frontières d'un mécanisme. Il s'agit essentiellement d'un critère pour déterminer si l'évidence expérimentale suffit à prouver qu'une corrélation entre phénomènes est due à une relation causale explicative. Ce critère s'inspire de la pratique expérimentale en sciences spéciales (biologie, psychologie, etc.). Ainsi, pour affirmer qu'une relation est causale, il faut vérifier (1) qu'agir sur le phénomène de haut niveau affecte de façon pertinente le phénomène de bas niveau, et (2) qu'agir sur le phénomène de bas niveau affecte de façon pertinente le phénomène de haut niveau. Par exemple, si on pense qu'un groupe de neurones contrôle certaines capacités motrices, on peut (1) amener l'organisme à utiliser ces dernières par une tâche quelconque et mesurer l'activité du groupe de neurones, et (2) stimuler ou inhiber les neurones en question et voir si cela affecte la capacité motrice qui nous intéresse.

Si deux phénomènes sont causalement liés, et qu'ils le sont de façon constante de telle sorte que l'un explique l'autre, alors on peut dire qu'ils forment un mécanisme. Dès lors, nous dit Kaplan, si on peut lier ainsi un phénomène de haut niveau qu'on considère traditionnellement comme cognitif avec un phénomène de bas niveau externe, alors on a montré qu'il y a extension cognitive.

Afin d'illustrer comment cette méthode peut aider à déterminer la frontière du cognitif, Kaplan emploie deux exemples tirés de la littérature sur l'esprit étendu. D'abord, il cite un exemple de Clark (1998) sur l'exploitation que fait le thon du courant et des propriétés hydrodynamiques de l'eau dans ses déplacements : en plus de détecter et exploiter les courants marins, celui-ci crée des gradients de pression et des vortex avec sa queue qui lui permettent d'aller plus vite que sa musculature ne le lui permettrait normalement. Altérer ces gradients et vortex (1) affecterait la vitesse du thon, de même que (2) faire nager le thon causerait

l'apparition de gradients et vortex exploitables : les conditions de manipulabilité mutuelle sont donc remplies, de sorte que les gradients et vortex font partie du mécanisme de locomotion du thon. Le deuxième exemple est celui de l'expérience de pensée d'Otto et Inga (Clark et Chalmers, 1998). Lorsqu'on fournit à Otto une information digne d'être mémorisée, comme une adresse, il la note dans son carnet et peut l'utiliser par la suite. Dès lors, si (1) on enlevait son carnet à Otto, ou (2) qu'on changeait l'adresse du MoMA, il serait incapable de s'y rendre : encore une fois, les conditions sont remplies pour montrer que la capacité d'Otto de mémoriser une adresse s'explique en partie par son calepin, de sorte que celui-ci fait vraiment partie du mécanisme de sa mémoire.

Une des critiques que Kaplan fait aux autres critères du cognitif qui militent pour une large extension, et en particulier à ceux de Clark et Chalmers (1998) et de Clark (2008), est le problème du grain. Comme nous l'avons mentionné dans la section précédente, Clark n'arrive pas à déterminer le degré de précision nécessaire pour faire la description d'un processus cognitif. Kaplan ne dit pas explicitement ce qui permet à la manipulabilité mutuelle de régler ce problème, mais laisse deviner que ça ait à voir avec sa capacité à discriminer les phénomènes pertinents à l'explication des conditions ambiantes (*background conditions*). En effet, si on s'attend à ce que des changements à une condition ambiante affectent un phénomène étudié, il n'est pas plausible qu'en agissant sur le phénomène, on modifie les conditions ambiantes. Kaplan prend pour exemple la relation entre la gravité et le contrôle sensorimoteur humain : Fisk *et al.* (1993) ont montré que le premier a un effet sur le second, mais, bien sûr, la gravité de la Terre ne peut être altérée par un humain. Si on prend un exemple dans le cerveau, on peut penser que la présence de myéline sur les axones est une condition ambiante pour la cognition : si on détruit les cellules qui la produisent (par exemple, par l'action d'un virus, comme c'est le cas chez les patients atteints de leucoencéphalopathie multifocale progressive), la cognition est définitivement altérée ; mais si on altère notre processus cognitif, la myélination ne semble pas affectée. Or la production de myéline est une fonction, et une fonction essentielle au maintien des neurones et à la propagation des potentiels – elle entre dans la réalisation d'autres fonctions de plus haut niveau, comme l'enregistrement d'une information dans la mémoire.

Elle fait donc partie du patron fonctionnel de la mémorisation, mais seulement si celui-ci est décrit avec un grain très petit. Ainsi, le problème du grain peut être résolu en considérant les détails des patrons fonctionnels à petit grain comme des phénomènes de bas niveau par rapport aux patrons fonctionnels de haut niveau.

Cependant, une telle méthode demande déjà que l'on définisse un patron fonctionnel à gros grain. Doit-on choisir l'évènement de mémorisation ? Ou devrait-on aller plus généralement à la fonction qu'est la mémoire ? Voire même la cognition en général ? Le critère de manipulabilité mutuelle ne permet de discriminer les détails fonctionnels pertinents de ceux qui ne le sont pas seulement dans la mesure où le phénomène de haut niveau, ou en tout cas le phénomène à expliquer, est défini. En ce sens, la méthode que propose Kaplan pose les mêmes problèmes que le principe de parité : il y a certes une heuristique de démarcation, mais le critère ne peut servir à caractériser le cognitif – pour cela, il faut encore une autre théorie.

On pourrait penser que la manipulabilité mutuelle permet néanmoins de changer les règles du jeu : comme elle offre un moyen de lier un phénomène de bas niveau à un phénomène de haut niveau dans un même mécanisme, elle pourrait permettre une défense d'une caractérisation de l'esprit exprimée dans un langage intentionnel. Cependant, comme on l'a dit au premier chapitre, la description intentionnelle dans ces caractérisations de l'esprit repose sur l'adoption d'une certaine stance interprétative de la part de celle ou celui qui produit cette description, de sorte qu'en dehors de quelques cas peu problématiques, il est difficile même de déterminer si un phénomène mérite une description intentionnelle.

Prenons le cas de l'intentionnalité. Comme l'attribution du caractère intentionnel à un phénomène relève de l'interprétation qu'on en fait, il est déjà très difficile d'imaginer une manipulation qui donne des résultats objectifs et valides pour l'application du critère de Kaplan. Cependant, cette interprétation dépend d'une stance qui précède l'expérience du phénomène – le caractère intentionnel d'un phénomène dépend du choix de l'interprète de le considérer comme tel. Il n'y a donc pas de manipulation pertinente possible pour un tel cas, et il s'en suit que

le critère de Kaplan est inapplicable.

Les autres marques du mental – l'accès privilégié, la conscience, etc. – ne font pas mieux à ce compte. C'est sans doute un peu parce que nos intuitions pour le mental et le cognitif eux-mêmes sont loin d'être claires – et qu'on peine à déterminer à partir du comportement si quelque chose est mental (ou cognitif) ou non. Mais c'est surtout que, lorsque ces intuitions interviennent, leur jugement ne porte pas sur une instance ou un évènement en particulier se produisant sur un moment, mais sur une capacité qui se manifeste sur une longue période, de sorte qu'elles ne sauraient opérer sur la base de quelques évènements isolés. L'intuition internaliste en est un bon exemple : du moment qu'on accorde qu'un humain a un esprit, elle dit que son cerveau contient du traitement cognitif, mais ne spécifie pas sous quelles conditions il se produit.

En somme, le principal problème du critère de Kaplan tient à ce qu'il n'est pas possible de formuler adéquatement le phénomène de haut niveau, à savoir le cognitif. En ce sens, comme le principe de parité, il s'agit d'un critère incomplet, qui attend encore une théorie du cognitif suffisamment précise pour permettre l'observation d'une manipulation.

2.4 L'accord avec l'évidence empirique

Un des aspects qui distingue le discours des philosophes qui cherchent à déterminer l'extension de l'esprit par rapport à celui de leurs prédécesseurs en philosophie de l'esprit, c'est l'importance que l'interprétation des travaux effectués dans les sciences empiriques et en informatique. Clark et Chalmers (1998) citent Kirsh et Maglio (1994); Kirsh (1995); McClelland *et al.* (1986); Hutchins (1995) afin d'illustrer comment l'usage d'objets dans l'environnement immédiat contribue cognitivement à certaines tâches, puis Suchman (1987); Beer (1990); Thelen et Smith (1996); Hutchins (1995) pour illustrer des domaines de recherches où la cognition est considérée comme continue entre l'organisme et l'environnement, et

enfin Ullman et Richards (1984); Blake et Yuille (1992) pour appuyer l'idée que nous exploitons les caractéristiques de notre corps afin de faciliter notre travail cognitif. Clark (2008) étend considérablement ce travail sur la littérature scientifique, citant au passage plus d'une centaine d'études et s'attardant sur plusieurs dizaines d'entre elles pour les commenter.

L'alternative à l'esprit étendu qui semble la plus prometteuse du point de vue empirique est la thèse de l'*esprit intégré* (*embedded mind*). Celle-ci est à vrai dire assez similaire à celle de l'esprit étendu. Contrairement à l'internalisme traditionnel, elle reconnaît pleinement le rôle fondamental de l'environnement dans les calculs du système cognitif (Rupert, 2009, 6). Elle reconnaît que l'esprit recrute des ressources extérieures pour s'économiser du travail cognitif, et reconnaît même l'existence de couplages avec l'environnement de l'organisme qui lui permettent d'accomplir certaines opérations cognitives. Cependant, pour elle, couplage ne veut pas dire constitution. Si un mécanisme accomplit un travail qu'on qualifierait de cognitif eût-il été dans le cerveau, on dira qu'il s'agit de traitement de l'information, mais une activité réellement cognitive doit s'accomplir dans le cerveau. En ce sens, la thèse de Rupert, bien qu'elle soit internaliste au niveau ontologique, est relativement similaire à l'esprit étendu.

Tous les philosophes s'intéressant à la question ne sont pas aussi portés que Clark à faire usage de l'évidence empirique pour appuyer leurs résultats, mais les partisans de l'internalisme – Rupert, Adams et Aizawa en particulier – bien qu'ils n'aient pas le zèle de Clark à cet égard, ne réclament pas moins avoir la recherche expérimentale de leur côté, et leurs prétentions ne sont pas sans fondement. Là où Clark mise sur l'analyse d'une grande variété d'études sans égard particulier à ce qui fait le plus grand consensus au sein des disciplines étudiées – après tout, l'esprit étendu est une thèse radicale, et il est normal que ses conséquences ne fassent pas consensus – Rupert tente de montrer que les résultats les plus robustes en psychologie cognitive supposent un esprit qui se réalise exclusivement dans le cerveau.

D'un certain angle, l'accord avec l'évidence empirique peut être considéré comme

une sorte de critère du cognitif – idée qui est explicitement mentionnée chez Barker (2010) et Rupert (2009), et qui semble implicite ailleurs. S’inspirant de concepts de philosophie des sciences, elle conçoit les différentes propositions d’extension du cognitif comme des théories scientifiques. Le rôle de l’évidence empirique est alors de départager entre ces différentes théories et de couronner celle qui s’accorde le mieux avec elle. Cette évaluation s’accompagne le plus souvent d’évaluations sur d’autres critères pertinents à l’évaluation des théories scientifiques, comme la simplicité. Cette évaluation globale des théories de l’extension de l’esprit en tant que théories scientifiques prend différentes formes selon les auteurs et les cadres qu’ils emploient, mais dans le cadre de celle-ci, l’évidence empirique a une place prépondérante, en particulier chez Clark et Rupert.

De fait, le seul critère de l’évidence empirique ne pourrait constituer un critère complet à lui seul, dans la mesure où il compte sur les théories en compétition pour déterminer les frontières. Cependant, ces dernières seraient souvent trop arbitraires indépendamment du contexte de compétition théorique dans lequel elles s’inscrivent. Le jugement de l’évidence empirique vient compléter la justification d’une théorie. Aussi, c’est peut-être l’ensemble que forment le jugement de l’évidence empirique et la théorie jugée gagnante qui peuvent former un critère complet au sens où nous l’avons défini plus haut.

2.4.1 Traitement de l’évidence empirique chez Clark

On peut légitimement critiquer le traitement que font l’un et l’autre aux travaux sur lesquels ils s’appuient. Ainsi, l’appui que Clark vient chercher dans les travaux empiriques est d’une nature un peu ambiguë. D’un côté, son *Supersizing the Mind* est essentiellement un exercice spéculatif : il s’agit de proposer une direction de recherche plus fructueuse. De l’autre, cette direction est présentée sous la forme d’une hypothèse, pour laquelle il argumente en présentant différents résultats, lesquels se veulent garants des succès futurs. Clark (2008, 115) présente donc le pouvoir de prédiction empirique de son hypothèse comme étant un *pari*, et le rôle

de la masse des travaux cités en appui à celle-ci est de justifier que ce pari est raisonnable, voire avisé.

Clark n'essaie donc pas de démontrer son hypothèse, mais plutôt de montrer qu'elle est plausible. Pour que l'évidence empirique suffise à soutenir cela, la barre est donc moins haute : il suffit de montrer que la spéculation s'appuie sur des hypothèses supportées par des observations. Cependant, même cette norme alléguée et plus indulgente demande quand même un minimum de rigueur. Il faut que les observations à partir desquelles on extrapole soient vérifiées, sans quoi elles ne peuvent être garantes de prédiction qu'on en dégage. Il faut donc éviter de spéculer à partir du fruit de spéculations, ou à partir des résultats d'études mal construites.

Cette façon plus libre de considérer l'évidence empirique n'empêche donc pas une comparaison des différentes théories de l'extension de l'esprit en regard de celle-ci. Clark (2008, chap. 6) l'illustre bien, alors qu'il commence par défendre que sa théorie s'accorde aussi bien avec les observations passées que celle de Rupert, avant de prédire qu'elle permet de prédire certains résultats à travers ses vertus « prophylactiques » (p.136-7).

Cependant la réussite de ce critère dépend de la qualité des observations qui sont employées comme support. Or il semble souvent que Clark ne remplisse pas toujours le critère minimal. Par exemple, Clark (2008, 133-4) dédie une page et demie à une simulation de Clowes et Morse (2005), laquelle est censée démontrer que le langage, ou plus précisément la possession de mots, nous permet d'optimiser notre apprentissage d'une façon qui nous donne un avantage évolutif. La méthode consiste à faire un agent virtuel fonctionnant avec un réseau de neurones, lequel évolue selon un algorithme génétique afin de maximiser sa performance dans un petit ensemble de tâches, et à permettre à l'agent de se renvoyer de l'information à lui-même à travers un canal spécial qui alimente l'entrée du réseau de neurones. Les chercheurs ont présenté trois conditions expérimentales : une sans ce canal, une avec le canal, et une avec le canal et une unité de sortie additionnelle qui permet de fermer et d'ouvrir le canal. La première condition a été la moins performante

aux tâches assignées, alors que la dernière a été un peu plus performante que la seconde. À partir de cela, et parce que les agents de la dernière condition avaient tendance à fermer davantage le canal dans les premières itérations, les chercheurs en ont conclu que lorsque l'agent acquiert une certaine connaissance de ses tâches, il se met à utiliser le canal de la même façon qu'on pourrait utiliser les mots pour structurer notre propre expérience.

Cette étude a plusieurs défaillances, qui militent pour qu'on la considère comme exploratoire, et non pas comme une évidence empirique robuste. D'une part, l'effet des conditions manipulées n'est pas correctement mesuré. Clowes et Morse ont peut-être fait plusieurs essais, mais une seule simulation est présentée, ce qui est un peu gênant dans la mesure où les résultats fluctuent considérablement au fil des générations. Ce l'est d'autant plus qu'aucun outil statistique n'a été utilisé pour évaluer les résultats : aucune hypothèse, nulle ou autre, n'a été formalisée, il n'y a pas non plus de calcul pour quantifier l'effet du hasard, et la différence entre conditions n'est quantifiée nulle part, de sorte que les auteurs en sont réduits à estimer les effets sur la seule base d'un histogramme de niveaux d'adaptation (*fitness level*). D'autre part, il n'est pas clair que la simulation soit appropriée pour représenter le travail structurant du langage sur le développement d'un agent qui le possède. Premièrement, le canal n'a même pas la permanence du mot, et ressemble à ce titre davantage à un simple retour d'information comme on en trouve beaucoup dans le cerveau, par exemple pour filtrer le bruit dans un signal ou pour faire ressortir des informations pertinentes – l'effet est tellement immédiat qu'on n'oserait même pas qualifier ce mécanisme de « mémoire ». Deuxièmement, même si on concédait que c'était une mémoire, ça ne serait pas pour autant un concept – lequel opère selon une certaine logique qui lui permet d'être pertinent sur toutes sortes de situations inconnues. Troisièmement, même si on arrivait à fermer les yeux sur cette difficulté, le langage est public et social – la difficulté particulière de celui-ci pour l'organisme consiste précisément à s'approprier quelque chose de conçu par quelqu'un d'autre dans des circonstances souvent différentes des siennes, et c'est précisément l'aspect que cette simulation ignore. Enfin, il existe plusieurs interprétations possibles pour les résultats obtenus par les chercheurs – par exemple, le seul fait d'avoir plus de contrôle sur ses interactions pourrait

suffire à expliquer les succès des agents des deux dernières conditions.

Or, dans un terrain qui conviendrait également à des spéculations allant dans tous les sens, non seulement Clark ne juge-t-il pas opportun de montrer les limites entre les conclusions que nous permettent les résultats et la spéculation à laquelle ils peuvent donner lieu, mais il en rajoute une couche. S’inspirant d’autres spéculations sur le fonctionnement des GasNet (Philippides *et al.*, 2005), il dit :

The power of verbal rehearsal to aid cognition might thus be explained as another instance of the more general value of autonomous, loose couplings between dynamically distinct processes. Perhaps, that is to say (and here we pick up some of the themes first introduced in chap. 3) self-produced verbal outputs enter into loosely coupled forms of coordination dynamics with nonverbal neural processes, allowing the overall system to explore trajectories through “thinking space” that might otherwise be blocked by destructive interference among superficially conflicting current ideas, goals, or contexts. (Clark, 2008, 134-5)

Pourtant, l’étude de Philippides *et al.* (2005) est elle aussi exploratoire – les auteurs qualifient eux-mêmes ses expériences de « préliminaires ». Elle propose deux variations des GasNets traditionnels inspirées de découvertes en neurologie, et montre qu’elles donnent de bien meilleures performances que les GasNets traditionnels. Pour expliquer ces résultats, les auteurs émettent l’hypothèse que des couplages flexibles (*loose couplings*) permettent d’augmenter les performances d’un réseau de neurones artificiel, et procèdent à différentes mesures pour vérifier cette hypothèse. Par cela, ils entendent une forme de couplage entre deux sous-systèmes qui permet une certaine coopération mais qui empêche que l’activité de l’un interfère avec celle de l’autre. Si les résultats de ces mesures sont compatibles avec l’hypothèse des auteurs, ils ne suffisent pas à la démontrer – en fait, il n’est même pas démontré que les variations réalisent vraiment des couplages flexibles.

L’objectif poursuivi par Clark dans cette section (§6.11) n’est d’ailleurs pas tout à fait limpide, mais il semble chercher à produire de l’évidence en faveur de cette proposition, qui apparaît dans la section suivante :

4. Most human cognizing benefits from cycles of self-stimulating acti-

vity (“cognitive turbo-drives”) in which we actively create the structures that drive and constrain our own evolving thought process. (Clark, 2008, 136)

Les sections précédentes ont déjà apporté de l’eau au moulin de cette hypothèse en analysant les travaux de Goldin-Meadow (2005) et de McNeill (2005) sur les gestuelles qui accompagnent la parole (*gestures*). Cependant, la discussion révèle que si les travaux dans ce domaine semblent montrer que les gestuelles accomplissent un certain travail cognitif, on ignore toujours quel est exactement l’apport cognitif des gestuelles. Aussi la section §6.11 semble-t-elle être un effort de produire une évidence convergente dans le domaine des réseaux de neurones. Elle tenterait ainsi lier les travaux sur les gestuelles et le travail de Clowes et Morse dans une hypothèse sur le genre d’avantages que peuvent avoir des recrutements comme ceux qui ont lieu dans ces études. Les « cycles d’activité autostimulatrice » formeraient donc un concept général rassemblant l’auto-stimulation du réseau de neurones de Clowes et Morse, les couplages flexibles des GasNets et l’activité auto-stimulatrice qui, croit-on, explique la contribution des gestuelles à la performance cognitive.

On doit se méfier de cette stratégie à plusieurs égards. Premièrement, autant l’auto-stimulation dans le réseau de neurones et dans les gestuelles que les couplages flexibles sont avancés dans le cadre d’hypothèses faites pour expliquer des phénomènes qu’on ne comprend pas très bien, lesquelles ne sont pas elles-mêmes corroborées. Deuxièmement, dans le cas des GasNets et du réseau de neurones, ces phénomènes eux-mêmes apparaissent dans des études préliminaires, l’une desquelles a de sérieuses défaillances. Troisièmement, Clark n’explique pas comment des mécanismes aussi différents que les trois types de mécanismes qu’il tente de subsumer sous le label d’« activité auto-stimulatrice » peuvent même légitimement être rassemblés ensemble.

En somme, la section §6.11 n’est probablement pas d’intérêt vital pour le livre ou même l’argument général du chapitre 6, mais elle ne rencontre certainement ses objectifs. Certes, le traitement que fait Clark à l’évidence empirique dans cette section relève peut-être de l’exception ; il s’agit peut-être d’un cas extrême qui se noie dans la mer de citations scientifiques de *Supersizing the mind*. Notre choix est

arbitraire, et une évaluation précise du traitement des évidences empiriques par Clark serait souhaitable pour en venir à une conclusion. Cependant, cet exemple nous montre que Clark n'est pas toujours minutieux dans ce genre d'exercice², et qu'il peut être vulnérable à une critique sur la fiabilité de ses conclusions.

Dans le chapitre 6, en réponse à Rupert qui en prend l'initiative, Clark emploie explicitement le critère de l'évidence empirique pour comparer les deux théories – aussi les recensions de travaux scientifiques peuvent-elles être interprétées dans ce cadre. Cependant, sa façon de l'employer l'amène à élargir ce qui compte comme évidence empirique pour prendre en compte les observations qui sont encore à faire. De fait, comme il en vient préalablement à la conclusion que l'évidence disponible ne permet pas de départager les deux théories, cet élargissement apparaît comme tout à fait sensé – il semble utile de faire parler l'évidence empirique au-delà de ce qu'elle dit, afin de voir apparaître les grandes tendances et de faire des prédictions qui peuvent nous aider à déterminer un gagnant.

Or, dans l'exemple qu'on a étudié, Clark extrapole à partir d'extrapolations. Un peu de réserve est toujours de mise quand on se trouve en présence de prédictions spéculatives – on présume que ces prédictions ont une certaine chance de faillir. Ce risque de faillite n'est peut-être pas calculable, mais un spécialiste du domaine peut en avoir une idée intuitive. Cependant, on peut se demander si on est toujours bien en mesure d'évaluer le risque d'échec lorsqu'on fait dépendre une spéculation d'une autre spéculation. Si la section §6.11 est garante du reste, alors Clark n'a peut-être pas d'évidence empirique future pour démarquer sa théorie, et on n'a donc pas non plus critère complet du cognitif.

2. Pour être honnête, la faute n'est certainement pas seulement celle de Clark, car la communauté philosophique en général semble avoir de la difficulté à évaluer critiquement des travaux qui ne relèvent pas de sa discipline. À témoin, Rupert, Adams et Aizawa eux-mêmes n'ont pas évalué le traitement des données scientifiques par Clark et par les autres partisans de l'esprit étendu. L'argumentation de la section §6.11 a été publiée deux ans plus tôt (Clark, 2006), mais Clark ne semble pas avoir reçu de commentaire à ce propos dont on puisse voir les répercussions dans le livre.

2.4.2 Traitement de l'évidence empirique chez Rupert

Rupert, quant à lui, consacre beaucoup moins d'espace de discussion à l'étude de l'évidence empirique, mais le fait sur une toute autre base, en imposant certaines exigences. Critiquant la tendance des défenseurs de l'esprit étendu à présenter des études isolées, qui portent sur un type particulier de phénomènes dans un type particulier de conditions, et à spéculer à partir de celles-ci, il suggère que la pratique en sciences cognitives est plutôt de viser à produire des résultats robustes. Par cela, il entend des résultats qui suivent des régularités d'une étude à l'autre. Ainsi, la *robustesse* des résultats revient à trois choses :

- (a) consistency in results as the concrete experimental design varies ;
- (b) consistent correlation between variations in results and variations in abstract properties of the materials used, a correlation insensitive to crosscutting variations in the materials used ; or (c) regularity in the changes of patterns of response (where concrete materials used can vary at each stage) as a function of past experience or age of the organism. (Rupert, 2009, 38)

D'emblée, le langage qu'emploie Rupert comporte certaines ambiguïtés. Si on la prend à la lettre, la condition (c) semble absolument impossible à atteindre : si seuls l'âge et l'expérience passée de l'organisme importent, alors on devrait avoir des résultats toujours semblables avec des sujets psychonormaux et des sujets ayant des handicaps cognitifs sévères. Par ailleurs, qu'entend-on par « propriétés abstraites des matériaux » ? On peut spéculer à partir des exemples que Rupert donne un peu plus loin que ce sont celles qui sont pertinentes à la tâche étudiée³. Ceci dit, il n'en reste pas moins que l'interprétation est incertaine. La façon la plus charitable de comprendre l'esprit de la définition est donc peut-être de n'accepter que les travaux où le changement dans les conditions ambiantes influe peu ou n'influe pas sur les résultats.

3. Par exemple, nous dit-il, pour la tâche de la fausse croyance, le taux de succès des enfants selon l'âge reste le même qu'on emploie des marionnettes, un conte oral, des découpages, des acteurs ou tout autre moyen de communiquer l'histoire (Rupert, 2009, 39). Rupert omet cependant de mentionner que la performance change selon ce qu'on prend comme indicateur d'une croyance (e.g. Onishi et Baillargeon, 2005).

Par ailleurs, il n'est pas tout à fait évident que ce genre de robustesse soit requise pour que des données empiriques soient prises en compte pour établir la caractérisation de l'esprit. Et quoique la consistance dans les résultats soit un bon indicateur de la qualité des données et de la crédibilité des hypothèses qui sont employées pour les expliquer, il n'est pas certain que ce soit toujours une condition nécessaire pour prétendre à ce genre de qualité. Par exemple, lorsque l'objet de la recherche est de mesurer l'effet de l'environnement sur un comportement, il va de soi que les conditions ambiantes influencent les résultats. Dans d'autres cas, l'objet d'étude est un phénomène complexe qui varie considérablement selon les conditions ambiantes, mais dont l'effet est mesurable et se confirme dans de nombreuses études. De tels travaux sont utiles pour notre compréhension du système cognitif, et bien que Rupert ne soit pas prêt à les considérer comme robustes, ils le sont souvent sous d'autres définitions en vertu de la constance avec laquelle leurs observations sont répétées.

Par exemple, Tversky et Kahneman (1973) relatent un ensemble de biais de jugement et de décision qui sont dus à ce qu'ils nomment « heuristique de disponibilité » (*availability heuristic*) : les stratégies qu'empruntent les sujets pour répondre à une question de l'expérimentateur vont souvent accorder plus d'importance aux informations qui sont faciles d'accès (autrement dit : « si je peux y penser, c'est que c'est important »). Ces heuristiques sont évidemment influencées par toutes sortes de facteurs : l'expérience passée, la trace laissée dans la mémoire, la capacité d'imagination et la forme de présentation (Tversky et Kahneman, 1973), l'amorçage (*priming* – Schwarz *et al.*, 1991), etc. De plus, l'heuristique de disponibilité prend souvent différentes formes, parce qu'elle s'adapte au problème auquel le sujet est confronté, de sorte que les facteurs qui influencent sa performance changent eux aussi. Pourtant, il semble absurde de considérer que les travaux qui portent là-dessus sont de moindre qualité en vertu de ces caractéristiques, ou qu'il s'agit d'un phénomène qui n'est pas à proprement parler cognitif. Le phénomène est tout simplement trop complexe pour rencontrer les exigences de robustesse de Rupert.

Non seulement le critère de robustesse de Rupert paraît-il trop restrictif, mais dans l’optique selon laquelle ce serait l’évidence empirique qui devrait déterminer la caractérisation et l’extension de l’esprit, il semble qu’il pipe les dés à un tel point qu’on voit mal comment n’importe quelle étude favorisant l’esprit étendu pourrait être considérée. Puisqu’il exige une constance dans une grande variété d’environnements et de contextes, alors il faut que les travaux qui rencontrent ces exigences étudient des processus qui se réalisent avec le plus petit dénominateur commun. Autrement dit, les études qui se penchent sur des phénomènes cognitifs et qui comptent beaucoup sur l’apport de l’environnement vont être d’emblée rejetées, puisqu’elles ne seront pas reproductibles dans les environnements qui ne peuvent fournir l’apport essentiel à la tâche que l’on demande.

Certes, dans certains cas, l’organisme s’adapte à divers environnements de façon à faire remplir le même rôle à des objets différents. Pour retrouver son chemin dans le bois, on peut marquer des arbres, mais dans le Nord, où il n’y a pas de végétation, on bâtit des Inuksuit. On peut donc imaginer que, pour une tâche, le sujet puisse décharger un travail cognitif particulier de diverses façons dans divers environnements, de telle façon à produire des résultats robustes dans le sens de Rupert. En ce sens, le critère de Rupert ne restreint peut-être pas le domaine empirique au point de rendre impossible toute preuve en faveur de l’esprit étendu. Cependant, la restriction rend tout de même le terrain plus difficile pour l’esprit étendu, de sorte qu’on aurait espéré une meilleure justification de sa part.

2.4.3 Empiriquement indistinguable

Entre les excès de confiance de Clark envers ses sources et la prudence excessive – voire idéologiquement biaisée – de Rupert, il y a peut-être un juste milieu où se réconcilient circonspection et justice épistémique. Cependant, encore faut-il que les théories qu’ils défendent se prêtent au test de l’expérience et soient proprement falsifiables. Or, selon Barker (2010) et Sprevak (2010), ce n’est probablement pas le cas : si on oppose l’hypothèse de l’esprit étendu telle que défendue par Clark

à celle de l'esprit intégré (*embedded*) telle que défendue par Rupert, le pouvoir explicatif est comparable, au point où il est impossible d'établir un vainqueur.

L'approche de Sprevak est de tester la prétention des deux auteurs quant à leurs hypothèses respectives – à savoir qu'elles représentent une inférence à la meilleure explication. Il passe en revue certains des principaux arguments émis de part et d'autre, et en vient à la conclusion que l'esprit intégré et l'esprit étendu ne font qu'échanger le poids de la complexité explicative entre théorie et ontologie⁴ : laisser l'esprit s'étendre dans l'environnement réduit l'ontologie, mais le réduire au cerveau rend la théorie plus économe. Difficile, alors, de dire ce qui vaut le mieux, nous dit Sprevak : on ne peut prédire les avantages de l'une ou de l'autre option.

Barker étudie plutôt le pouvoir de prédiction des deux théories, et en vient lui aussi à la conclusion qu'elles prédisent exactement les mêmes comportements. Il en déduit qu'elles doivent aussi avoir la même complexité, comprise en termes de nombre de paramètres ajustables. Barker reconnaît que certaines prétentions de l'esprit étendu, à savoir qu'il posséderait un avantage pragmatique qui éviterait certaines erreurs et ferait avancer davantage la psychologie, peuvent avoir certains mérites. Cependant, il reconnaît lui aussi qu'il n'est pas en mesure de tester cette affirmation.

Pour l'un comme pour l'autre, (1) les théories sont indistinguables empiriquement, (2) il est improbable qu'une expérience survienne qui nous permette de les départager, et (3) les deux théories s'équivalent en termes de simplicité. Cependant, l'impact que peuvent avoir les différentes ontologies de l'esprit sur l'élaboration de la recherche scientifique reste un enjeu. Prenons par exemple l'impact qu'elles ont dans la dynamique d'évolution de la recherche que décrit Lakatos (1970).

Pour Lakatos, la recherche scientifique se fait dans le cadre de *programmes de recherche*, qui rassemblent des théories scientifiques qui partagent des intuitions

4. On entend ici « ontologie » dans un sens qui est peut-être plus proche de l'usage qu'en font les informaticiens que de celui qu'en font certains philosophes : c'est, en gros, la façon que l'on a de produire un découpage du réel.

communes. Celles-ci partagent aussi des concepts et donc des ontologies communes, et sont cohérentes les unes avec les autres. Au centre des programmes de recherche se trouve un *noyau dur* de théories qui incarnent les intuitions communes. Ces théories sont en ce sens essentielles au programme de recherche, de sorte que lorsque l'évidence empirique semble les contredire, les chercheurs tendent à tout faire pour éviter d'avoir à les modifier. Les théories du programme de recherche qui ne font pas partie du noyau dur font partie de la périphérie, et servent à deux choses : d'une part, elles se présentent comme l'extension logique de leur programme de recherche, et permettent à ce dernier d'augmenter son domaine de pertinence, et d'autre part, on peut en les modifiant expliquer des données empiriques allant à l'encontre des prédictions des théories du noyau dur sans pour autant toucher à ce noyau dur.

Lakatos ne parle pas explicitement des ontologies, mais parle des énoncés « *syntactiquement métaphysiques* », qui n'ont pas de conditions de vérité empiriques, et qui ne peuvent donc pas être réfutés par une expérience. Une définition est un tel énoncé : bien qu'il n'y ait aucun cheval sur Terre né avec une corne sur le front, on ne considère pas que la définition de la licorne est falsifiée. Elle est toujours valide, même si elle n'a aucune extension. Aussi l'ontologie d'un domaine appartient-elle à ce domaine des énoncés syntactiquement métaphysiques.

Pour Lakatos, bien qu'ils ne peuvent être réfutés par une expérience, ces énoncés peuvent l'être par une autre théorie. Par exemple, prenons la définition aristotélicienne de la vie : est vivant tout être qui se meut ou croît de par soi-même (la croissance étant conçue comme une forme de mouvement). Les connaissances que nous avons acquises depuis l'antiquité ne réfutent pas cette définition – elle produit toujours des catégories tangibles. Cependant, il existe d'autres définitions. Par exemple, prenons la définition qu'on retrouve couramment dans les manuels de biologie, selon laquelle est vivante toute entité qui consomme et se reproduit : celle-là, en revanche, est contradictoire avec la définition aristotélicienne. La réfutation d'une définition survient de la même façon qu'avec n'importe quelle théorie : à savoir quand le programme de recherche auquel elle contribue est remplacé par un autre parce qu'il n'arrive plus à produire de nouveau contenu empirique (sous

forme de nouvelles applications théoriques produisant de nouvelles prédictions ou sous forme de prédictions confirmées par l'expérience empirique).

Autrement dit, l'essor d'une ontologie dépend des programmes de recherche qu'elle permet. Si ces programmes de recherche s'avèrent couronnés de succès, alors l'ontologie se voit elle aussi validée, autant parce qu'elle reste pertinente que parce qu'il est révélé qu'elle peut permettre l'exploration scientifique.

Une telle évaluation prend du temps, puisqu'il faut pouvoir évaluer le développement des programmes de recherches sur de longues périodes pour voir l'impact de l'ontologie sur ceux-ci. Aussi pourrait-on suivre le conseil de Clark, soit « laisser un millier de fleurs s'épanouir », et attendre que le temps choisisse une ontologie et le programme de recherche qui lui correspond. Or, dans le cas qui nous intéresse, même si on s'accorde avec Clark, Sprevak et Barker pour dire que l'enjeu est substantiel, on n'est pas obligé de croire que les deux théories aboutissent nécessairement à des programmes de recherche différents. Certes, Clark et Rupert tendent à privilégier des recherches différentes : le premier promeut celles qui mettent en valeur l'apport de l'environnement dans la cognition, alors que le second met l'emphase sur les travaux de psychologie cognitive qui mettent de l'avant l'apport de l'organisme. Cependant, ces directions ne sont pas inconsistantes, et peuvent facilement être complémentaires. On peut facilement imaginer qu'elles participent toutes deux d'un programme de recherche « situationniste », qui tenterait de comprendre l'esprit humain dans le contexte de son environnement. Les deux directions de recherche renforceraient l'idée fondamentale du programme situationniste, à savoir que l'on doit comprendre la cognition à travers les relations que l'organisme entretient avec son environnement : par exemple, l'étude des processus cognitifs transcâniens nous donnerait une idée de ces relations, alors que l'étude des caractères pérennes de l'individu nous dirait comment l'individu adapte ces interactions à divers types d'environnements. Quant aux différences d'ontologie, si on peut s'accommoder d'alterner entre explications en termes de systèmes dynamiques et d'explications en termes d'interfaces et de traitement d'information, on peut probablement s'accommoder de l'alternance entre les ontologies de l'esprit étendu et de l'esprit intégré. Aussi, même si l'esprit intégré

et l'esprit étendu participaient réellement de programmes de recherche différents, ces programmes de recherches ne trouveraient pas suffisamment d'inconsistances entre eux pour qu'on puisse affirmer qu'ils sont en compétition.

À ce point-ci, on serait tentés de penser que le débat qui fait rage entre esprit étendu et esprit intégré est un faux débat, parce que dans la mesure où on peut les distinguer, ils ne semblent nullement en opposition. De fait, si le débat était purement et simplement une question d'ontologie scientifique, ce serait sans doute le cas. Cependant, le concept d'esprit correspond aussi à un concept commun à de nombreux domaines scientifiques et autres (dont les sciences sociales, comme nous l'avons évoqué au précédent chapitre). De plus, les détails de la caractérisation de l'esprit portent souvent une certaine puissance d'évocation qui peut influencer l'usage qu'on en fait dans chacun de ces domaines. Aussi, Clark (2008, 136) n'a-t-il pas tort de soulever la question des vertus prophylactiques de l'esprit étendu. Sans doute l'ontologie du concept d'esprit, étant donné le bagage culturel et théorique qui l'accompagne, est-elle en mesure d'influencer l'emploi qu'en font les chercheur-ses. Cependant, ce n'est pas seulement l'ontologie qui fait ce travail, mais l'ontologie avec le bagage en question. C'est donc le concept dans son sens large qui doit ici être étudié.

Une autre façon de poser le problème serait la suivante : si on considère que les cœurs du programme de recherche de l'esprit intégré et du programme de recherche de l'esprit étendu se limitent à l'ontologie, alors ces programmes sont équivalents ou tout au moins compatibles ; mais si le cœur est formé de l'ontologie et d'autres contenus théoriques relevant des concepts d'esprit qui s'associent à ces ontologies différentes, alors il est possible qu'ils soient effectivement contradictoires. Si tel est le cas, dépendant des concepts en question, il se peut même qu'il soit possible de trancher à partir de données empiriques. Ceci dit, si les cœurs se limitaient effectivement à l'ontologie, on voit mal comment ils pourraient produire à eux seuls les développements théoriques qu'on doit espérer d'un programme de recherche. Charitablement, on doit donc considérer que c'est la seconde option qui est la bonne : l'esprit étendu et l'esprit intégré se démarquent davantage que par leur seule ontologie.

On ne peut blâmer Sprevak et Barker d'avoir réduit la différence entre les deux théories à cette seule dimension ontologique, car cette abstraction contribue à clarifier les différences entre esprit intégré et esprit étendu. Cependant, ce faisant, on laisse de côté d'autres aspects pertinents à cette compréhension. Pour pouvoir terminer l'évaluation de l'accord des théories avec les données empiriques, il faudra d'abord mieux formaliser ces théories. En ce sens, on est encore renvoyé à un nouveau travail de théorisation.

2.5 Le contenu non-dérivé

Le dernier critère important qui renvoie effectivement à d'autres travaux de théorisation est celui de la représentation. Il se présente sous plusieurs formes, mais son intuition fondamentale vient de l'association que fait le cognitivisme entre cognition et représentation. Les auteurs auxquels cette idée est le plus volontiers associée sont Adams et Aizawa (2008, 2001). Adams et Aizawa défendent en fait un critère fait de deux conditions nécessaires. Selon eux, pour qu'un processus soit cognitif, il doit d'une part (1) impliquer des contenus non-dérivés et d'autre part (2) être réalisé par le bon type de processus.

La deuxième condition part du principe selon lequel il y a quelque chose dans le mécanisme des neurones qui ne se retrouve pas dans les autres tissus organiques qui leur permet de réaliser des processus cognitifs. Le principe a peut-être du mérite, mais Adams et Aizawa ne le développent pas davantage, ni pour apporter des pistes pour l'éclairer, ni pour suggérer des façons de le faire. Aussi est-il difficile de l'évaluer.

La première condition, en revanche, se base sur une intuition partagée depuis longtemps par les philosophes de l'esprit, et qui se retrouve même chez la majorité des externalistes (incluant Clark et Chalmers, 1998; Menary, 2007; Rowlands, 2010b; Wheeler, à paraître). Par contenu, Adams et Aizawa entendent le contenu sémantique ou le contenu d'une représentation; celui qui rentre dans les explications

cognitivistes du comportement. Lorsque le contenu d'une représentation dérive du contenu d'une autre représentation préexistante, alors il est dérivé : c'est le cas d'un panneau de circulation, que l'on interprète comme ayant un sens en vertu de notre compréhension préexistante du code de la route. Aussi, le panneau en soi (indépendamment des normes par lesquelles on l'interprète) ne possède pas le contenu qu'on lui attribue généralement ; il ne veut pas dire grand chose s'il n'y a personne pour le lire, et sa signification idoine n'apparaît pas non plus si la personne qui le regarde ne maîtrise pas les conventions nécessaires pour l'interpréter. L'existence de contenu dérivé laisse croire qu'il devrait y avoir du contenu non-dérivé, qui n'a pas besoin d'autre représentation pour faire sens. Ce type de contenu, croient Adams et Aizawa, doit provenir de notre esprit, et est donc un bon candidat pour être une marque du cognitif, c'est-à-dire une condition nécessaire à ce que quelque chose soit cognitif.

La conception de la représentation comme contenu non-dérivé est singulièrement intentionnelle, et à ce titre, il est difficile de se donner une idée d'où elle pourrait se réaliser, comme on l'a déjà mentionné au premier chapitre et dans la section §2.2. En ce sens, il est difficile de comprendre comment ce critère peut être éclairant pour déterminer l'extension ou la localisation du cognitif. Pour répondre à cette question, Adams et Aizawa (2008, 36-7) proposent trois façons de déterminer si une structure ou un mécanisme réalise du contenu non-dérivé, en faisant appel à diverses théories appartenant au domaine de la sémantique naturalisée (Dretske, 1981, 1988; Fodor, 1989, 1990; Cummins, 1996). Mais plutôt que de trancher et d'évaluer si le modèle choisi réalise bien les représentations là où ils les attendent, c'est-à-dire dans le cerveau, Adams et Aizawa ne font que conclure platement que les modèles tentent « évidemment » de montrer comment le contenu émerge d'une matière qui, *a priori*, n'en a pas.

Une des choses les plus étranges de l'argumentation des deux auteurs est que les deux opinions qu'ils épousent explicitement à propos du contenu non-dérivé, à savoir qu'il est condition nécessaire du cognitif et qu'il se retrouve dans le cerveau, ne sont pas appuyés par une argumentation positive tant que par une critique des opinions contraires. Le but semble être de montrer que c'est une possibilité, mais

il ne semble pas y avoir de soucis de prouver que c'est une théorie plausible.

La raison pour cette stratégie se trouve peut-être dans une vision assez pessimiste de notre compréhension de l'esprit :

We can read biology textbooks that explain how acids and enzymes in the mouth, stomach, and intestines cause the breakdown of certain molecules in food, but we have yet to find a chapter in a neuroscience textbook that provides a comparable explanation of how brains generate non-derived representations. (Adams et Aizawa, 2008, 36)

Dans les faits, il existe certes des travaux qui proposent des explications quant à la façon de faire émerger des représentations internes dans des organismes. Par exemple, à l'aide de simulations (Cangelosi et Harnad, 2001) et de travaux sur les robots (Steels, 1999; Vogt, 2000), on peut produire des conditions qui permettent de faire émerger des concepts que des agents artificiels peuvent s'échanger pour mieux répondre aux tâches qui leurs sont demandées. Cependant, Adams et Aizawa semblent penser que ces travaux ne sont pas encore mûrs pour donner une idée générale de la production de contenu non-dérivé tel qu'on l'entend communément, c'est-à-dire celui qui se produit dans nos cerveaux. Si le couperet de l'expérience n'est pas tombé, toute théorie est acceptable – et particulièrement celle pour laquelle penchent peut-être le plus volontiers nos intuitions (Adams et Aizawa, 2001, 33-4).

Cependant, Adams et Aizawa comprennent que l'intuition ne saurait suffire pour trancher le débat. Aussi attendent-ils le couperet de l'expérience. Cependant, il y a dans cette attente l'admission que les sémantiques naturalisées, telles qu'on les connaît, ne suffisent pas à décrire l'émergence de la représentation. Il n'est pas impossible que les auteurs croient que le problème tient du manque de consensus (ils y font allusion : Adams et Aizawa, 2008, 55), mais *pro tanto*, les sémantiques naturalisées semblent plutôt donner des conditions nécessaires à l'émergence de contenu que de véritables explications. Aussi les auteurs déplorent-ils que ces conditions puissent être interprétées de plusieurs manières : en d'autres mots, on peut difficilement passer de la théorie à l'implémentation, et dans cette mesure, les critères d'Adams et Aizawa faillissent donc à produire une caractérisation claire.

Les auteurs réclament donc explicitement une théorie plus complète de sémantique naturalisée pour trancher le débat. En somme, la condition de contenu non-dérivé telle que l'utilisent Adams et Aizawa tombe clairement dans la catégorie des critères qui en appellent à davantage de théorie.

2.6 Conclusion

Dans ce chapitre, nous avons étudié des critères parmi ceux qui sont le plus souvent avancés comme critères du cognitif. De fait, certains d'entre eux sont discutés au point d'être emblématiques des discussions sur l'esprit étendu.

Le principe de parité, même s'il n'a pas la prétention d'être un critère, est souvent exploité comme tel. En explorant ses variations, nous avons montré qu'il ne saurait en tout cas être considéré comme un critère complet, et que sa principale contribution est plutôt de souligner qu'un critère du cognitif doit être *a priori* agnostique quant à la localisation des processus. Le critère de manipulabilité mutuelle, quant à lui, peut être un outil très utile, mais il laisse lui aussi en suspens la notion de cognitif. Le critère d'accord avec l'évidence empirique repose quant à lui sur les prédictions que font les diverses hypothèses sur la notion de cognitif en terme de conséquences empiriquement mesurables. Cependant, la formulation de ces hypothèses a elle-même besoin d'être étoffée au-delà de la simple dimension ontologique, sans quoi la comparaison sur le plan empirique devient un exercice trivial et absurde. Enfin, le critère de présence de contenu non-dérivé ne produit pas de caractérisation de ce qui constitue un contenu non-dérivé en dehors du langage intentionnel.

Dans ces quatre cas, l'incomplétude du critère nous empêche de déterminer s'il nous amène vers une frontière du cognitif. Dans le cas du principe de parité et de la manipulabilité mutuelle, c'est assez évident, puisque la notion de cognitif est elle-même laissée en suspens. Dans le cas des hypothèses qui entrent en jeu dans la comparaison sur le plan empirique, puisqu'on ne peut trancher sur la meilleure

ontologie, on ne peut déterminer quelle hypothèse propose la meilleure extension du cognitif. Enfin, impossible aussi de trancher avec le critère d'Adams et Aizawa, puisqu'il ne se prononce pas sur la façon de localiser les représentations.

De plus, il est difficile de dire si l'intuition individualiste est à l'œuvre. Dans le cas du principe de parité et de la manipulabilité mutuelle, encore une fois, ça semble évident. Sans notion du cognitif, on pourrait adopter presque n'importe quelle perspective. On peut peut-être évaluer les conceptions de l'esprit étendu de Clark ou de l'esprit intégré de Rupert sur cet aspect, mais si on réduit le programme aux ontologies, alors on ne peut dire s'ils se basent sur l'intuition individualiste. Enfin, c'est probablement le cas qu'Adams et Aizawa sont influencés par l'intuition individualiste, mais tel qu'énoncé, leur critère de contenu non-dérivé ne se mouille pas assez pour se baser sur elle.

Ces critères ne sont donc pas directement utiles à l'évaluation de notre hypothèse, à savoir que si un critère dépeint le cognitif comme réalisé dans un espace qui possède des frontières. Nous pouvons donc passer à la prochaine partie : celle des critères « complets », qui sont, eux, susceptibles de nous aider dans notre investigation. Ce sont eux qui forment le véritable corpus d'étude, sur lequel nous testerons notre hypothèse.

CHAPITRE III

INDIVIDUALISME

Nous arrivons enfin au nœud de notre argumentation. Comme nous l'avons dit plus haut, ce chapitre et le précédent servent à établir l'hypothèse H2, selon laquelle les critères du cognitif qui déterminent son extension à l'aide d'une frontière reposent sur l'intuition individualiste. Pour ce faire, nous avons choisi de passer par l'analyse de texte. Le chapitre précédent portait sur la construction du corpus d'étude, le réduisant aux critères dits « complets », parce qu'ils permettent de déterminer une extension du cognitif. Dans le présent chapitre, nous clarifierons l'hypothèse afin d'en tirer des prédictions sur le corpus d'étude, lesquelles serviront à faire un test qui nous aidera à déterminer si l'hypothèse s'avère.

3.1 Méthodologie

Pour chaque critère étudié, nous opérerons en deux étapes : une étape d'interprétation, suivie d'une étape plus spéculative, où nous manipulerons les prémisses de l'auteur·e afin de voir l'impact sur les conclusions. La première étape sert à établir la présence d'une association entre des indicateurs de l'intuition individualiste et des indicateurs de frontières de l'esprit. La seconde étape sert quant à elle à vérifier qu'il ne s'agit pas d'un effet du hasard ou d'un phénomène externe, mais que cette association résulte des liens conceptuels entre les deux thèses.

Dans la première étape, nous tenterons de déterminer, d'une part, si l'intuition individualiste peut être détectée chez l'auteur·e du critère étudié, et, d'autre part, s'il y a des indications que le critère donne lieu à des frontières du cognitif. En répondant à ces deux questions, nous pourrions évaluer s'il y a une association entre l'une et l'autre.

Comme il s'agit d'un sujet complexe qui demande une discussion approfondie, il sera question en section 3.2 de l'intuition individualiste, et de la façon de la reconnaître lorsqu'elle est implicite dans un texte. Concernant la présence de frontière, le concept est certainement plus simple, mais il n'est pas sans complexité : que dire, par exemple, des contours flous, des formes discontinues, ou des frontières qui évoluent dans le temps selon l'activité de l'organisme ? Notre objectif ici est d'abord de faire sauter ce qui nous semble être un préjugé sans fondement ; aussi n'y a-t-il pas de raison de ne pas être généreux dans notre définition de frontière, en autant que ce mot n'en perde pas son sens et qu'on reste dans le même préjugé. Pour ce faire, il faut cependant une définition de frontière qui soit susceptible de découler de l'intuition individualiste. C'est pourquoi nous clarifierons aussi ce que nous entendons par « frontières » et notre démarche pour les reconnaître dans la section 3.2.

La seconde étape s'inspire de la méthode de Craver (2007b,a) pour établir une relation de causalité entre deux variables. De fait, que l'intuition individualiste soit présente chez les mêmes auteur·es qui conçoivent le cognitif comme ceint par des frontières ne dit pas grand chose sur la relation entre l'une et l'autre. Sans rentrer dans le détail de cette relation, on peut tester un de ses aspects : on peut vérifier si, dans le cadre de la réflexion d'un·e auteur·e, l'adoption de la première thèse est liée causalement à l'adoption de l'autre.

Pour ce faire, on suppose que ce lien causal existe : dès lors, si l'auteur croit que le cognitif doit être ceint par une frontière (thèse F), c'est parce qu'il souscrit à l'intuition individualiste (thèse I). Formellement, on a donc $F \Rightarrow I$ puisque l'on suppose qu'il ne peut y avoir d'autre raison pour justifier F (on suppose que c'est vrai dans le cadre de la littérature étudiée), et on a $I \Rightarrow F$ puisque l'hypothèse que

l'on veut vérifier est que de souscrire à I nous amène nécessairement à souscrire à F lorsqu'on prend position sur l'extension du cognitif dans le cadre de la littérature étudiée. Si tel est le cas, et si on suppose que l'auteur·e devait ne plus souscrire à I , alors ille devrait être amené·e à ne plus souscrire à F , puisque la justification de F repose sur I . Inversement, si on imagine un scénario où l'auteur·e ne souscrit pas à F , alors le contexte de sa pensée, il devrait jaillir une certaine contradiction à ne pas souscrire à F .

Dès lors, s'il y en effectuant ces deux manipulations à l'aide de raisonnements contre-factuels, on peut tester notre hypothèse (formellement : $F \Leftrightarrow I$).

3.2 Intuition individualiste et frontières

À la base, ce que nous entendons ici comme étant l'intuition internaliste découle de deux suppositions fondamentales. La première concerne l'esprit lui-même, qui est considéré comme une substance indivisible, synthétique et distincte du non-cognitif. Ces caractéristiques correspondent à l'expérience qu'on fait de notre propre vie cognitive : une continuité et une consistance dans notre conscience comme dans nos actes, une dichotomie sujet-objet au niveau de l'action et de la perception. Autrement dit, la phénoménologie de notre vie cognitive nous la représente comme appartenant à un individu et comme le constituant. La seconde concerne le rapport entre l'esprit et ce qui le réalise, ou entre le niveau personnel et le niveau sous-personnel. En substance, cette supposition veut que notre connaissance de l'esprit au niveau personnel puisse nous informer sur ce qui le réalise au niveau sous-personnel, de telle sorte qu'à partir de la structure générale de l'esprit tel qu'on le connaît au niveau personnel, on puisse émettre des contraintes sur ce qui peut le réaliser. En particulier, on suppose qu'il y a un certain degré d'isomorphie entre niveaux. Si l'on combine ces deux suppositions, on est tentés d'appliquer au niveau sous-personnel l'individualisme que l'on retrouve au niveau personnel : les mécanismes qui réalisent l'esprit appartiendraient donc à un individu, et le réaliseraient.

La présence de frontières du cognitif découlerait aussi de cette translation : en effet, la qualité d'appartenir à et de réaliser un individu est quelque chose qu'un mécanisme peut posséder ou peut ne pas posséder – il n'y a pas d'entre-deux. Aussi doit-il y avoir une frontière. Et comme l'individu est indivisible et synthétique, cette frontière doit entourer un espace fermé et sans discontinuité. La paroi qui sépare le cognitif de ce qui ne l'est pas devrait donc produire quelque chose comme un solide.

3.2.1 La critique du schéma entrée-sortie

Les deux suppositions derrière l'intuition individualiste ont, à notre sens, un certain mérite, en ce sens qu'elles ne doivent pas être rejetées *a priori*, parce qu'elles répondent à certaines intuitions que l'on peut avoir à propos de l'esprit, et valent donc dans la mesure où elles apparaissent comme plausibles. Cependant, elles doivent être traitées comme des suppositions qui doivent être vérifiées et critiquées. Ainsi, on peut remettre en question l'idée d'isomorphie entre le personnel et le sous-personnel sur la base du fait qu'il apparaît comme un concept particulièrement vague : à quel niveau d'abstraction au niveau personnel et au niveau sous-personnel décrit-on la forme où ils devraient se rejoindre ? Autrement dit, quelles sont les formes et les caractéristiques pertinentes à cette isomorphie ?

Ce genre de remise en question n'est pas inédit. À titre d'exemple, dans *Consciousness in Action* (2002), Susan Hurley attaque une type représentation de l'esprit qui emploie ce qu'elle appelle le « schéma entrée-sortie » (*Input-Output Picture*). Pour expliquer la nature de cette conception, il faut comprendre qu'au niveau personnel, une personne est renseignée sur le monde en le *percevant* et qu'elle l'affecte en *agissant* sur lui ; alors qu'au niveau personnel, nos sens reçoivent des données de l'extérieur (des *entrées*) et notre système cognitif émet des signaux à nos muscles (des *sorties*). Le schéma entrée-sortie confond ces deux dichotomies, et en vient ainsi à comprendre la perception comme réception d'entrée, et l'action comme l'émission de signaux aux muscles (Hurley, 2002, 287).

C'est une erreur, nous dit Hurley, en grande mesure parce que la perception et l'action émergent en réalité des couplages entre l'organisme et le monde, qui font que la perception comprend une part de motricité de la part de l'organisme, et que l'action comprend une part sensitive encore plus importante.

Pour le démontrer, elle fait appel à des expériences de pensée qui exploitent un type de scepticisme qui est le produit de la confusion du schéma entrée-sortie. Dans un tel schéma, les interfaces sont claires, et peuvent être considérées en un bloc (perceptions ou actions) selon leur direction. Aussi, puisqu'elles peuvent être considérées en bloc, elle peuvent donc aussi être remises en question en bloc : si on ne fait sens du monde externe que sur la base d'un flot de données, comment peut-on savoir si l'interprétation que l'on fait de cette information est juste ? Les images qu'elles provoquent en nous pourraient bien n'être qu'une fabrication élaborée. De même, si l'action n'est qu'une question de sortie, qu'est-ce qui nous garantit que notre tentative de poser un geste va donner le geste que l'on souhaite ? La perception et la tentative dans le schéma entrée-sortie sont des interfaces entre l'esprit et le monde. En ce sens, ce schéma nous amène à considérer l'esprit comme relativement indépendant du monde, et cette indépendance est la condition qui donne lieu au scepticisme.

À partir du scepticisme se développent les expériences de pensée de « duplication » : ces cas où une personne est téléportée dans un environnement identique au sien à l'exception de quelques différences fondamentales (e.g. Block, 1990; Putnam, 1973). Dans ces expériences, on s'arrange généralement pour que le sujet ne soit pas conscient des changements dans son environnement : si le spectre des couleurs est inversé, on lui donne des verres de contact qui le réinversent ; si le monde est inversé comme dans un miroir de sorte que tout ce qui se trouvait à gauche du sujet se trouve à droite, et inversement, on lui donne des lunettes qui refont l'inversion, etc. L'idée est de tester nos intuitions pour voir si le contenu de sa perception, ou l'intention dans sa tentative d'agir, sont identiques à ce qu'ils étaient dans leur environnement original.

L'exercice fonctionne plutôt bien si on ne fait qu'inverser le spectre des couleurs.

Cependant, du moment qu'une modification à l'environnement prend une dimension spatiale – et donc, du moment que nos capacités motrices peuvent être affectées – on a besoin de mécanismes de compensation complexes pour corriger nos mouvements, ce qui nous force à corriger notre proprioception, et ainsi de suite. Chaque correction affecte un autre couplage, et en appelle ainsi à de nouvelles corrections, et celles-ci se rapprochent ainsi de plus en plus des fonctions cognitives centrales – du coup, le sujet n'est plus si identique dans son nouvel environnement qu'on pourrait le souhaiter. À moins, bien sûr, qu'on veuille faire reculer la frontière de ce que l'on considère comme « interne » à chaque fois que l'on doit faire une correction plus intrusive. Mais d'une part, la frontière risquerait alors de varier selon l'expérience de pensée choisie, et d'autre part, il ne resterait souvent plus grand-chose d'interne.

Hurley utilise cette analyse des expériences de pensée comme une *reductio ad absurdum* contre le schéma entrée-sortie. D'une part, il n'est pas possible de réussir ces expériences de pensée parce qu'il n'y a pas de frontière où l'information ne va que dans un sens : la perception et l'action se caractérisent au niveau sous-personnel par des couplages où l'information va dans les deux sens. Comme les modalités par lesquelles l'humain interagit avec ce qui l'entoure échangent de l'information dans les deux sens, le schéma entrée-sortie n'est pas approprié pour le décrire. D'autre part, la faillite du schéma entrée-sortie ouvre la possibilité de l'esprit étendu. En effet, l'absence d'interface où l'information ne va que dans un sens nous prive de ce qui aurait été la frontière entre environnement et esprit, ce qui ouvre déjà la possibilité que les deux domaines se chevauchent. Comme cela signifie la présence de couplages entre le cerveau, le corps et l'environnement, la contribution de ces derniers au travail cognitif est plus importante, ce qui ouvre la possibilité que des mécanismes s'étendant sur ces trois médiums puissent constituer l'activité cognitive de l'esprit.

3.2.2 L'insaisissable légèreté du lien entre niveaux

La faillite du schéma entrée-sortie nous apprend par ailleurs cette leçon : lorsqu'il est question de l'esprit, toute supposition que l'on fait concernant l'isomorphie entre niveaux doit être employée avec la plus grande prudence. Parce que les descriptions de haut niveau sont souvent émergentes¹ par rapport aux descriptions de bas niveau, c'est trivialement vrai pour tout système complexe, et ce l'est d'autant plus pour un système aussi élaboré que le système cognitif humain.

Cette leçon ne tient pas que pour l'isomorphie : prédire une caractéristique à un niveau à partir d'une caractéristique à un autre niveau relève toujours de la spéculation, sauf lorsque l'on dispose de données empiriques précises qui établissent un lien de causalité entre elles. Par exemple, on dit souvent de la dopamine qu'elle est le neurotransmetteur du plaisir et de la récompense, parce qu'elle joue un rôle important dans des phénomènes comme l'apprentissage, la dépendance (à l'alcool, aux drogues, etc.) et l'orgasme. Cependant, augmenter la quantité de dopamine dans le cerveau en administrant son précurseur L-DOPA à des humains ne semble pas affecter leur quantité ou leur qualité de plaisir en général (Liggins *et al.*, 2012) : le mécanisme est beaucoup plus compliqué, et la présence de dopamine ne peut être assimilée au plaisir.

Si l'isomorphie (ou tout autre extrapolation du haut vers le bas niveau) peut servir d'inspiration – si elle peut participer de certaines logiques qui permettent de faire des prédictions sur les résultats que produiront certaines avenues de recherche expérimentale – toute spéculation qui provient de ce procédé demande à être vérifiée (sous forme l'établissement d'une relation causale) et expliquée (sous forme d'un mécanisme de réalisation).

Cette leçon s'applique aussi à la relation qui lie l'intuition individualiste – entendu comme l'intuition de haut niveau – aux caractéristiques qu'on en induit au niveau sous-personnel – la frontière du cognitif. Si des philosophes en viennent à présumer de la présence de frontières à cause de l'intuition individualiste, leur présomption

1. Au sens de Wimsatt (2007).

est donc vulnérable au même genre de critique que celle qui est dirigée envers le schéma entrée-sortie.

Certes, il est tentant d'utiliser des informations qu'on possède à un niveau de description pour produire par extrapolation des prédictions sur le niveau inférieur. Ceci dit, pour l'isomorphie, la leçon provenant de critiques comme celle de Hurley semble avoir été apprise. Ainsi, à l'exception d'Adams et Aizawa (2005), personne dont le nom est associé au débat sur l'extension de l'esprit n'emploie explicitement ce principe, et beaucoup (incluant Clark et Chalmers, 1998) le récusent carrément. On ne peut que supposer que les critiques à son endroit ont eu leur effet. Par ailleurs, aucun autre principe qui permettrait de déduire des caractéristiques de niveau sous-personnel à partir de caractéristiques de niveau personnel n'est proposé.

Aussi, supposant que l'hypothèse que nous tentons de démontrer est vraie, il faut que les philosophes concernés aient fait ce passage du personnel au sous-personnel de façon implicite, voire même à leur insu. Les philosophes sont, comme tout le monde, sujets aux biais implicites. De plus, toute pensée, philosophique ou autre, repose ultimement sur des croyances non-justifiées, voire non-examinées. Cette situation est donc plausible. Ceci dit, que le lien entre intuition individualiste et frontière reste implicite rend notre travail particulièrement délicat : il faut pouvoir l'établir en dépit du silence qui l'entoure.

3.2.3 Frontières

Comme le lien entre intuition individualiste et frontière, s'il existe, reste implicite, il est assez difficile de déterminer ce en quoi il consiste. On peut néanmoins spéculer un peu sur la nature de ce lien.

Notre hypothèse est qu'il s'agit d'un lien d'isomorphie, ce type d'opération nous apparaissant comme la raison la plus plausible. À vrai dire, juger de la nature

exacte de l'opération qui sous-tend cette relation est délicat parce que l'opération de pensée derrière l'application du principe d'isomorphie n'est pas un raisonnement logique, mais un raisonnement analogique, dont l'issue est ouverte. Ainsi, une caractéristique de l'intuition individualiste peut-elle se manifester dans divers aspects de la description de niveau sous-personnel du système cognitif – par exemple, ce qu'un philosophe voit se manifester dans l'extension peut se manifester dans la forme temporelle des opérations pour un autre. Aussi est-il impossible de formaliser le processus de pensée. En revanche, on peut faire certaines conjonctures en posant le problème de la bonne façon.

Prenons la première caractéristique : l'intuition individualiste dicte que quelque chose appartient (constitutivement) à l'individu ou non, ce qui amène une extension discrète. La bivalence du haut niveau pourrait se manifester de plusieurs façons au niveau sous-personnel. Cependant, la question se pose différemment si notre travail n'est pas de déterminer les conséquences de l'intuition individualiste (ce que les philosophes ne font vraisemblablement pas si elle reste implicite), mais plutôt de prédire la forme de l'extension. Aussi, dans le cas de la forme de la frontière, la question se pose-t-elle plutôt comme suit : « Étant donné l'intuition individualiste, est-il plus probable que la frontière soit discrète, ou non ? ». Or les philosophes qui débattent de l'extension de l'esprit sont plutôt intéressés à cette dernière question.

Posé en ces termes, procéder par isomorphie semble invitant. On a là un scénario idéal pour appliquer une heuristique de substitution d'attribut, que Kahneman et Frederick (2002) définissent ainsi :

We will say that judgment is mediated by a heuristic when an individual assesses a specified *target attribute* of a judgment object by substituting another property of that object – the *heuristic attribute* – which comes more readily to mind. (p. 53)

L'attribut ciblé (*target attribute*) est ici la forme de la frontière, et l'attribut heuristique celle de l'individu au niveau personnel, tel que nous le transmet l'intuition individualiste. Épistémologiquement parlant, la difficulté ici réside en ceci qu'on peut difficilement manipuler ou même mesurer la présence de l'attribut heuris-

tique pour vérifier que c'est bien cette heuristique qui est à l'œuvre. Cependant, si l'on suppose que l'intuition individualiste est présente et disponible pour un-e philosophe, alors, en l'absence d'autre information pertinente, il semble plausible que cette personne ait recours à l'heuristique de substitution d'attribut.

Suivant ce scénario, si le problème est de savoir si la frontière est discrète ou non, on substituera la frontière par l'individu tel que posé par l'intuition individualiste. Et comme l'intuition individualiste dépeint une dichotomie nette entre l'agent et le monde, la réponse sera que la frontière du cognitif doit être discrète.

La situation est analogue pour la seconde caractéristique. La question que l'on se pose est alors : « L'extension du cognitif est-elle continue, ou y a-t-il des discontinuités ? » Encore une fois, l'intuition individualiste suggérant un agent continu, toujours par substitution d'attribut, on en vient à décider que la frontière doit entourer une région continue.

Le résultat est donc une frontière discrète ceignant un espace sans discontinuité – bref, une frontière qui approxime un solide géométrique. Ces attributs suffisent à caractériser ce qu'on entend par « frontière » dans le cadre de notre investigation. Aussi la méthode que nous emploierons pour savoir si un critère du cognitif détermine une extension ceinte par une frontière consistera à chercher pour ceux-ci. Ou, dit autrement : si l'extension peut être approximée par un solide, alors il y aura frontière.

3.2.4 Détecter l'intuition individualiste

Le traitement que nous avons fait jusqu'ici de l'intuition individualiste se limitait aux aspects qui touchent les frontières – autrement dit, aux aspects se traduisant au niveau sous-personnel sous la forme de frontières. Aussi, si on limitait à ces seuls aspects pour notre analyse de texte, alors les heuristiques pour détecter les frontières et l'intuition individualiste feraient appel aux mêmes marques

observables dans le texte. Si celles-ci indiquent à la fois la présence de l'intuition individualiste et la présence de frontière, alors on ne pourra détecter l'une sans l'autre – dans un tel scénario, l'hypothèse H2 serait nécessairement vraie. Cette dernière serait ainsi triviale, et l'étude du corpus deviendrait inutile. Il nous faut un moyen de déterminer la présence d'une des deux notions dans la pensée d'un·e auteur·e de façon à ce que celle-ci soit *a priori* indépendante de la présence de l'autre concept. Il nous faut donc élargir notre perspective à son égard.

Prenons par exemple l'autonomie du cognitif. Une impression que nous donne notre expérience de notre propre vie cognitive est que nos représentations nous appartiennent en propre, de même que les opérations de la pensée dans lesquelles elles sont impliquées. Et comme on peut au moins en produire certaines de notre propre gré, il semble que l'agent soit à l'origine de ces phénomènes cognitifs. Aussi en vient-on à supposer que notre système cognitif est à la fois ce qui les produit et les réalise – en ce sens, il est autonome.

Le schéma entrée-sortie garantit cette autonomie à l'aide de la frontière particulière qu'il suppose : si la perception n'est qu'une affaire d'entrées, alors les représentations (ou du moins, ce qui les véhicule) doivent se trouver à l'intérieur des frontières, puisque l'on doit être capable de les manipuler. De même, comme on peut planifier l'action, l'intention qui l'anime doit se trouver elle aussi à l'intérieur, puisque l'interface de l'action n'est qu'une affaire de sorties, et que l'on doit pouvoir manipuler nos intentions. Cependant, la disparition du schéma entrée-sortie ne signifie pas pour autant que l'on renonce complètement à l'autonomie du cognitif : des frontières où l'information va dans les deux sens sont tout à fait compatibles avec celle-ci – simplement, elles ne la garantissent pas.

Un signe que l'autonomie est bien indissociable d'autres aspects de l'intuition individualiste est que l'autonomie et le concept de frontières du cognitif (tel qu'on doit le concevoir suite à la critique de Hurley) entretiennent une relation symbiotique. D'une part, l'autonomie donne une nouvelle motivation pour une frontière discrète et continue : celle-ci permet d'en déterminer l'environnement par rapport auquel les processus cognitifs peuvent être autonomes. D'autre part, elle garan-

tit les conditions empiriques nécessaires pour que l'établissement d'une frontière discrète et continue soit possible. En effet, pour qu'on puisse tracer une frontière, il faut qu'il y ait un espace à l'intérieur duquel se constituent entièrement les processus cognitifs – autrement dit, un espace à l'intérieur duquel le système cognitif est entièrement constitué. Or un tel espace est impossible à ceindre si des couplages l'étendent indéfiniment. L'autonomie des processus prédit que leur réalisation s'avérera distincte du monde dont ils sont autonomes, et garantit ainsi qu'un tel espace délimité est possible.

En revanche, la présence de frontières ne permettant plus de garantir l'autonomie du cognitif, celle-ci doit s'expliquer d'autres façons. En particulier, l'autonomie des opérations suppose toujours une autonomie des représentations en ce sens qu'elles sont manipulées par l'agent, et ce même si leur contenu se réalise en dehors de celui-ci. Parallèlement, même si les moyens d'action peuvent dépendre constitutivement de ce qui est extérieur à l'agent, l'initiative de celles-ci doit encore lui revenir. Aussi, dans le genre d'autonomie qui est supposé par l'intuition individualisme, a-t-on de bonnes chances de voir apparaître une hiérarchie d'une sorte et/ou de l'autre, puisque celle-ci permet d'expliquer qu'un mécanisme appartient ou non à l'agent : si celui-ci est dépendant, pour son sens, d'une faculté centrale représentant l'agent, alors il est cognitif ; sinon, il ne l'est pas².

La présence d'une hiérarchie à l'intérieur du système cognitif peut facilement être détectée dans la pensée d'un·e auteur·e, puisqu'elle est souvent explicite. Aussi peut-elle être considérée comme un signe de la présence de l'intuition individualiste. Bien entendu, d'autres motivations pourraient expliquer la présence de cette hiérarchie, et on pourrait sans doute adopter l'intuition individualiste sans que ça se traduise nécessairement dans une hiérarchie des facultés à l'intérieur du système cognitif. Cependant, si on ne peut trouver une autre motivation à la présence d'une telle hiérarchie, on a un signe que l'intuition individualiste peut être à l'œuvre.

Une autre caractéristique qui peut nous aider à détecter l'intuition individualiste

2. Dans les sections suivantes, on verra certains exemples concrets de cette hiérarchie.

dans un texte est la présence d'un principe d'individuation. Au niveau personnel, la question de l'unité de la personne fait encore l'objet de nombreux débats. Cependant, l'unité individuelle à travers le temps, telle qu'elle nous apparaît communément dans notre expérience, est généralement considérée comme un phénomène à expliquer, et non comme un phénomène à établir. Au niveau sous-personnel, cependant, l'individualité du système cognitif ne devrait pas aller de soi : rien n'indique que ce caractère de l'expérience phénoménale doive nécessairement se transférer au mécanisme qui la réalise. Cependant, les conditions sont tout aussi favorables à un transfert du caractère individuel de l'esprit vers le niveau sous-personnel qu'elles le sont pour le transfert de son caractère discret et séparé du monde. Aussi peut-on s'attendre à ce que les philosophes qui souscrivent à l'intuition individualiste tiennent pour acquise l'unité individuelle à travers le temps et l'espace. Et comme, même si l'on ferme les yeux sur les difficultés que pose ce transfert du niveau personnel au niveau sous-personnel, cette unité demande toujours à être expliquée, on peut s'attendre à trouver des tentatives de le faire.

Pour notre tâche de détection de l'intuition individualiste, le transfert du caractère individuel du système cognitif est plus ou moins important : la seule valeur ajoutée par rapport au caractère de continuité et de discrétude que nous donnent déjà les frontières est une correspondance à l'individu au niveau personnel – la personne. En revanche, le questionnement parallèle de ce qui peut être à l'origine de cette unité (conçue comme une réalité contingente) est particulier. L'explication des frontières se trouve dans la caractérisation de l'esprit : on doit trouver là ce qui justifie qu'on trace la frontière à un endroit. En revanche, l'explication du caractère individuel doit nous dire pourquoi le système cognitif a pris cette forme individuelle plutôt qu'une autre. C'est donc cette caractéristique qui nous intéressera, et qui sera le second signe de la présence de l'intuition individualiste.

Enfin, l'individu, au niveau personnel, possède des droits et des devoirs en vertu de la place qu'il occupe dans notre compréhension de l'éthique et du droit. Aussi les états et processus mentaux au niveau sous-personnel en viennent-ils à hériter des égards qu'on doit à la réalisation physique d'une personne. Par exemple, Clark et Chalmers (1998) mentionnent qu'une atteinte à un objet qui appartient

au système cognitif étendu d'un individu devient, dans une perspective d'esprit étendu, une atteinte à l'intégrité de la personne – ce qui peut justifier des peines particulièrement sévères, par exemple, pour le vol d'un iPhone. Cela peut amener à repenser les frontières du public et du privé. Ainsi, si la critique de Hurley permet de séparer frontières de l'esprit et frontières de l'environnement, et ainsi permettre à l'esprit d'empiéter sur l'environnement, la partie de ce dernier sur laquelle le cognitif aempiété est non seulement devenue esprit, elle est devenue un espace privé, appartenant en propre à celui qui l'a envahi. Du coup, si on entend l'environnement comme un espace public, l'esprit ne se retrouve plus tout à fait dans l'environnement – il l'a annexé, et l'a soumis à ses propres droits.

Il est difficile de penser ce transfert des droits individuels aux états et processus sous-personnels sans supposer l'intuition individualiste – et en particulier sans faire l'adéquation entre le système cognitif au niveau sous-personnel et la personne au niveau personnel. Justifier des droits semblables sans supposer qu'il s'agit d'une même essence demanderait une justification additionnelle, laquelle ne se retrouve pas dans la littérature qui nous intéresse. On peut donc considérer que, lorsqu'un-e auteur-e soulève la question du statut moral d'une partie du système cognitif, c'est un signe qu'elle ou il souscrit à l'intuition individualiste.

Bref, il semble qu'on ait trois indices pour reconnaître la présence de l'intuition individualiste dans un texte (ou un ensemble de textes) : (1) si on retrouve une forme de hiérarchie à l'intérieur même du système cognitif entre un lieu central et sa périphérie, (2) s'il y a un souci de trouver au niveau sous-personnel le principe d'une individuation robuste et transversale, à l'image de celle qui se trouve au niveau personnel et (3) si les états et processus cognitifs au niveau sous-personnel héritent des droits qui sont accordés à l'individu au niveau personnel. Inversement, s'il n'y a dans l'énonciation d'un critère du cognitif aucune de ces caractéristiques, ou que celle qui s'y trouve peut s'expliquer par d'autres raisons, alors on dira que l'intuition individualiste n'est pas supposée.

Ces indices forment ensemble une heuristique, laquelle ne saurait donner de résultat certain pour un texte ou ensemble de textes individuellement, mais peut nous

permettre d'appuyer notre hypothèse lorsqu'appliquée sur un plus large échantillon. Nous l'appliquerons donc sur les quatre critères que nous avons jugé remplir les conditions que nous avons énoncées en section 2.1 : Rupert (2009), Clark (2008), Haugeland (1993) et Rowlands (2010b).

3.3 Rupert

Comme l'internalisme est la voie traditionnelle et qu'elle était plus ou moins incontestée par les cognitivistes et les philosophes de l'esprit, ceux qui la défendent ne ressentent souvent pas le besoin de formuler un critère pour la justifier, et se contentent de critiquer son alternative. Rob Rupert fait exception. Afin d'appuyer l'option qu'il soutient (l'esprit intégré), il propose comme critère ce qu'il appelle « *System-Based View* »³ (Rupert, 2009, 41-4). Le nom vient de son désir de distinguer ce qui est réellement cognitif en tant qu'espèce naturelle de ce qui ne fait que contribuer (même de façon importante) au cognitif.

Rupert décrit ainsi le fonctionnement de son critère :

Consider, then, a more precise characterization of what it is for a set of resources to be integrated. Each token instance of cognitive behavior (alternatively, each act of completing a cognitive task) in a given subject involves the causal contribution of certain mechanisms, abilities, or capacities, factors that make a causal contribution distinctively to the production of the cognitive explananda in question. Thus, for a given subject at a given time, there exists a set of mechanisms, capacities, external resources, etc. each of which has contributed distinctively to that subject's cognitive processing on at least one occasion. For each such type of mechanism, relative to each kind of cognitive phenomenon that it helps to produce, there is a conditional probability (determined by relative frequency, if by nothing else) of its use relative to each of the other mechanisms, abilities, etc. in the set, as well as a conditional probability of its use relative to each subset the-

3. L'auteur choisit ici de ne pas traduire cette expression, qui est déjà passablement galvaudée en anglais chez Rupert.

reof. These conditional probabilities can be rank ordered. Assume that mechanisms A , B , and C have contributed distinctively to cognitive processing in the subject in question at least once (but not necessarily to the same token process). Now assume that, given the history of the co-contribution of A , B , and C , $P(C|A\&B) = 0.7$; then the set $\{A, B, C\}$ goes on the list as a 0.7 (relative to some particular kind of cognitive outcome) and will likely come in ahead of many other sets (for a particular subject at a particular time). The same set might also appear at a different place in the ordering because, for instance, $P(A|B\&C)$ does not equal 0.7 relative to the same kind of cognitive outcome. Moreover, given the variety of kinds of cognitive outcome, the same sets appear on the list many more times, most likely with many different associated probabilities. Next, consider the likelihood of a natural cutoff between the higher probabilities and lower ones, a gap that separates highly co-contributing mechanisms from those that are less so. (In the absence of a significant gap in the ordering, 0.5 would seem to mark the relevant cutoff point.) Now count the number of times each type of mechanism appears on the list of sets with higher conditional probabilities (i.e., those sets above the significant gap on the list); these frequencies themselves exhibit a rank ordering, and a natural cutoff (another significant gap) separates those mechanisms that appear frequently on the list—that is, are highly interdependent and heavily co-employed—from those that appear rarely. This indicates which mechanisms are part of the integrated set to be identified with the cognitive system and which are, in contrast, resources used by the cognitive system. (p. 42-43)

En somme, disons que pour chaque mécanisme, il y a une probabilité d'avoir contribué causalement à un phénomène cognitif, et disons qu'il y a des mécanismes dont cette probabilité est affectée par le fait qu'un autre mécanisme contribue lui aussi au même phénomène cognitif – e.g. le mécanisme A a beaucoup plus de chance de causer un phénomène cognitif si le mécanisme B est aussi activé. Cela donnerait lieu à une probabilité conditionnelle – e.g. $\Pr(A | B)$ la probabilité que le mécanisme A contribue à un phénomène cognitif étant donné l'activation de B . Le pari de Rupert est qu'on trouvera un ensemble de mécanismes qui s'activent ensemble pour causer des phénomènes cognitifs, de sorte que leur probabilités conditionnelles seront fortes entre elles, mais plus faibles avec les autres mécanismes. Cet ensemble de mécanismes serait le système cognitif.

On pourrait ici penser que, n'ayant pas défini ce qui fait qu'un phénomène est cognitif, Rupert se trouve devant le même genre de problème que Kaplan (cf. section 2.3). Cependant, la supposition de Rupert est que les mêmes mécanismes sont responsables d'une très large proportion de phénomènes cognitifs. Dès lors, ce sont ces mécanismes, et non pas les phénomènes de haut niveau qui en découlent, qui définissent ce qui est cognitif. Donc pas besoin, comme chez Kaplan, de déterminer l'ensemble des tâches qui sont cognitives pour savoir quels mécanismes le sont aussi ; au contraire, ce sont les mécanismes qui, réalisant les tâches qui sont cognitives, permettent d'en déterminer le statut cognitif.

Cependant, la supposition de Rupert sur l'organisation des mécanismes du cognitif est un peu arbitraire, car celui-ci ne nous donne pas nécessairement beaucoup de raisons explicites de penser qu'elle doit s'avérer. Certes, il y a une volonté de rendre compte d'une certaine littérature en psychologie cognitive⁴. Cependant, comme on l'a dit plus tôt, le critère de robustesse introduit un biais en faveur de l'hypothèse de Rupert, parce que le sujet « robuste » est celui qui réagit similairement malgré les changements dans son environnement – un tel sujet pourrait donc difficilement avoir un système cognitif qui s'étend au-delà du corps. De plus, il n'est pas particulièrement approprié de fournir de l'évidence empirique venant de la psychologie cognitive pour appuyer une supposition qui se situe au niveau sous-personnel : elle peut certes nous permettre d'affirmer des choses sur les phénomènes cognitifs, mais beaucoup plus difficilement sur les mécanismes qui les réalisent.

Il est donc raisonnable de penser que l'origine de cette supposition puisse être l'intuition individualiste. D'une part, le transfert de la notion d'individu du niveau personnel au niveau sous-personnel est assez explicite. Rupert explique clairement que son objectif est de rendre compte de l'individu tel qu'il se manifeste dans les travaux de psychologie qui l'intéressent. D'ailleurs, il qualifie les alternatives qui souscrivent plus ou moins à l'esprit étendu de « no-self views », parce que l'explication de l'individu, juge-t-il, n'est pas un souci central dans ces théories.

4. Certes, Rupert s'éloigne parfois un peu de la psychologie cognitive (e.g. Rupert, 2009, 40), mais l'emphase est explicitement sur cette discipline en particulier.

On peut donc dire que le critère de Rupert nous donne explicitement ce qu'il faut pour reconnaître l'indice (2) : le souci de trouver au niveau sous-personnel le principe qui permet la formation de l'individu tel qu'on l'entend au niveau personnel. C'est d'ailleurs tout le sens du critère de Rupert : la co-contribution est une propriété vérifiable empiriquement et indépendante censée expliquer cette individualité.

Certes, les autres indices sont plutôt élusifs. Ainsi, on ne trouve pas vraiment de hiérarchie à l'intérieur de ce que Rupert classe comme étant le cognitif, et on ne trouve pas non plus de discussion sur le statut moral de ce dernier. Cependant, ces absences peuvent être, sinon expliquées, du moins justifiées. Concernant la hiérarchie, comme le critère de robustesse qu'emploie Rupert est assez restrictif, il semble qu'il garde les frontières du cognitif au plus proche du centre, de sorte qu'il ne reste plus beaucoup de place pour établir une hiérarchie à l'intérieur des frontières du cognitif. Concernant le statut moral, Rupert n'en parle pas vraiment, mais c'est probablement parce que la frontière qu'il propose est conservatrice. Comme l'idée que le cerveau est seul à avoir un statut cognitif particulier fait partie de l'orthodoxie, son statut cognitif a déjà été abordé dans les grandes lignes, et toute réflexion additionnelle à son sujet ne revient pas à la philosophie de l'esprit ou des sciences cognitives, mais à la bioéthique. Bref, on peut parier que Rupert accorde un statut moral particulier au cerveau, mais que ce n'est pas son rôle de l'aborder.

Le livre de Rupert n'exhibe donc qu'un indice de la présence de l'intuition individualiste, mais cet indice est clair, et Rupert est assez explicite quant au rôle essentiel de l'individu dans la mise sur pied de son critère. Pour lui, expliquer l'individu, la personne est le rôle fondamental de la psychologie cognitive, et le critère du cognitif doit refléter cette vision. Aussi semble-t-il difficile de nier la présence de l'intuition individualiste dans la pensée de Rupert.

Par ailleurs, Rupert est explicite quant à la localisation des mécanismes qui remplissent son critère : ils doivent se trouver bien intégrés et connectés entre eux, ce qui semble supposer qu'ils se réalisent dans le cerveau (Rupert, 2009, 53), où

les connexions sont à son avis plus nombreuses qu’avec l’extérieur de cet organe. De toute façon, étant donné cet impératif d’intégration, et le statut très clair des mécanismes – soit ils sont cognitifs, soit ils ne le sont pas – il est difficile de voir comment on pourrait échapper à une frontière du cognitif. En ce sens, le critère de Rupert remplit clairement les conditions pour la présence de frontières.

Qu’arrive-t-il si on manipule son argumentation ? Si on enlève l’intuition individualiste, la supposition de Rupert ne tient plus. Du coup, il n’est pas certain que l’on arrive à un seul ensemble de mécanismes co-contribuant aux phénomènes cognitifs. Si tel est le cas, les liens entre les différents ensembles peuvent être ténus – au point ils pourraient être assez éloignés les uns des autres. Dès lors, la forme géométrique la plus appropriée pour approximer l’extension ne serait pas un solide, mais bien plusieurs. Donc enlever l’intuition individualiste ne préserve pas nécessairement les frontières. Inversement, comme nous l’avons montré, les frontières sont une conséquence directe de l’intuition individualiste chez Rupert – dès lors, si on devait supposer une autre forme d’extension, on devrait abandonner l’intuition individualiste.

3.4 Clark

Il n’y a pas, chez Clark, de critère du cognitif à proprement parler, parce qu’une fonction cognitive peut se réaliser de fort diverses façons – il est plutôt un partisan du « *motley crew* » (se traduit littéralement par « équipage bigarré » – Clark, 2008, 95) – et s’analyser selon plusieurs cadres conceptuels – il propose d’ailleurs qu’il faille comprendre l’esprit autant comme un système dynamique qu’un organe de traitement d’information (Clark, 2008, 27).

Bien que, comme on l’a dit en section 2.2, il conçoive le principe de parité comme une pompe à intuition davantage que comme une théorie, Clark reste attaché au fonctionnalisme sur lequel il repose. Aussi le statut cognitif des processus dépend essentiellement de leur profil fonctionnel au niveau fonctionnel. Or, tout dans le

profil fonctionnel n'est pas pertinent pour déterminer le statut cognitif : Clark (2008, 89) argue que c'est davantage au niveau du « gros grain » (e.g. la capacité à retenir de l'information) qu'au niveau du « fin grain » (e.g. un profil fonctionnel détaillé de la mémoire, incluant le réseau synaptique) que se situe l'information pertinente. Si le détail fonctionnel peut varier au niveau du fin grain, alors toutes sortes de choses peuvent remplir une fonction cognitive, pourvu qu'elles répondent au patron fonctionnel plus général – d'où l'« équipage bigarré ».

Pour distinguer le cognitif de ce qui ne l'est pas, Clark propose donc plutôt quelque chose qui ressemble à une heuristique, qu'il applique sur une grande variété de cas tirés de la littérature en sciences cognitives. On part d'une idée intuitive du profil fonctionnel d'une tâche ou d'un processus cognitif – Clark adhère en effet à ce qu'il appelle, après Braddon-Mitchell et Jackson (2006), « fonctionnalisme de sens commun », selon lequel l'individu humain normal possède une riche théorie implicite des profils fonctionnels qu'épousent les tâches et différents processus mentaux de niveau personnel (Clark, 2008, 88). À travers l'étude de différents travaux en sciences cognitives, Clark tente de montrer qu'au niveau sous-personnel la réalisation de ces derniers implique un flot d'information qui forme un réseau de couplages – et donc un système dynamique. Ce flot d'information, quant à lui passe à travers diverses *composantes*, lesquelles transportent ou traitent l'information qui les traverse. Ces composantes possèdent des frontières, puisqu'elles possèdent aussi des interfaces auxquelles l'organisme peut les coupler ou les découpler à d'autres composantes pour opérer un traitement d'information. Le réseau de composantes couplées pour l'accomplissement de la tâche ou du processus sont la base matérielle qui la ou le réalise.

Diriger ce flot d'information et l'organiser dans un tout est cependant l'essentiel du problème ici. En ce sens, Clark n'est pas particulièrement original lorsqu'il mentionne que les processus cognitifs ne le sont que dans la mesure où ils participent au système cognitif dans son entier (Clark, 2008, 86). La solution de Clark à ce problème provient de ceci qu'il souscrit à ce qu'il appelle l'hypothèse de la cognition centrée sur l'organisme (*Hypothesis of Organism-Centered Cognition* ou *HOC*, Clark, 2008, chap. 6). Celle-ci veut que le principe de l'organisation du

cognitif réside tout de même dans l'organisme, et, pour l'humain, plus particulièrement dans son cerveau – en tout cas, dans ce qui constitue le centre névralgique de l'organisme. Le rôle de ce centre névralgique quant à cette organisation consiste entre autres choses à gérer le couplage et le découplage des différentes ressources cognitives selon les besoins de l'activité de l'organisme (Clark, 2008, 135) – en d'autres mots, il est responsable du recrutement des ressources cognitives, qu'elles soient ou non à l'intérieur du corps de l'organisme.

C'est ce qui permet à Clark de dire que les frontières de l'esprit sont toujours négociables au rythme des changements d'action et de recrutements (Clark, 2008, 34). En ce sens, il faut distinguer le concept de frontières qui apparaît dans la critique que Clark en fait (e.g. Clark, 2008, 73) de celui qu'il utilise lorsqu'il parle de frontières négociables. Comme Hurley, Clark est sceptique quant à la pertinence du schéma entrée-sortie, et considère l'idée qu'on se fait traditionnellement de la démarcation intérieur/extérieur comme arbitraire. Cependant, la solution, pour lui, reste de trouver une autre frontière – une frontière dynamique, certes, mais frontière néanmoins selon notre définition, en ceci que la forme qu'elle produit peut effectivement approximer un solide.

Clark semble adopter l'intuition individualiste malgré son opposition au schéma entrée-sortie. L'objectif des sciences cognitives, selon lui, est d'abord d'expliquer les performances cognitives de l'agent cognitif, qu'il identifie comme étant « a reliable, easily identifiable physical nexus of perception and action, apparently driven by a persisting and modestly integrated body of goals and knowledge ». L'explication, développe-t-il, peut certes nous amener à considérer des facteurs qu'on n'aurait initialement jamais imaginé – comme par exemple des contributions de l'environnement. Cependant, la supposition reste qu'il faut trouver au niveau sous-personnel ce qui correspond à l'individu au niveau personnel.

Aussi, bien que Clark insiste parfois sur la diversité des tactiques d'analyses, et que cela suppose effectivement différentes façons de découper le monde, son heuristique d'analyse semble vouloir donner des raisons indépendantes de rattacher les différentes composantes qui font le travail cognitif à l'individu pour lequel ce

travail est fait. Ainsi, il y a clairement chez Clark une hiérarchie à l'intérieur du système cognitif : entre, d'une part, le centre exécutif dans le cerveau qui gère l'organisation des couplages, qui est à l'origine du phénomène de recrutement de ressources cognitives hors cerveau (Clark, 2008, 135) et au-delà duquel ne s'étend pas la conscience (Clark, 2009), et, d'autre part, les ressources périphériques qui sont recrutées par ce système.

Par ailleurs, on note que cette organisation de l'esprit à travers le recrutement par le couplage est une hypothèse empirique (l'étude des réseaux de couplages devrait nous éclairer sur le sujet) et de surcroît d'une hypothèse sous-personnelle, puisque la tâche de gérer cette distribution se fait automatiquement, sans réflexion consciente. Si elle s'avère, elle garantit effectivement le concept d'une cognition centrée sur l'individu, et de la présence de frontières qui répondent à nos critères. Elle consiste donc en un principe sous-personnel qui garantit une unité qu'on devine au niveau personnel. Voilà donc un deuxième indice que Clark souscrit à l'intuition individualiste.

Enfin, Clark a été parmi les premiers à suggérer que l'esprit étendu avait des impacts au niveau moral (Clark et Chalmers, 1998). Une atteinte au carnet d'Otto, ou au iPhone de Chalmers (2008), est possiblement une atteinte à l'intégrité de la personne. Lorsque l'individu recrute quelque chose pour du travail cognitif, il envahit de ce fait la chose recrutée et l'investit des droits qui lui appartiennent. Clairement, Clark opère un transfert de l'individu du personnel au sous-personnel : nos indicateurs semblent donc signaler qu'il y a intuition individualiste.

Encore une fois, la manipulation maintient le lien entre intuition individualiste et frontières. On a montré que la présence de frontières était un résultat de l'adoption de l'intuition individualiste ; aussi est-il difficile de supposer un autre genre d'extension sans tomber dans la contradiction. Par exemple, si on imagine le genre d'extension qui conviendrait avec un système cognitif de groupe, on aurait du mal à la réconcilier avec l'idée d'un système central qui gère le réseau de couplages. Enfin, si on laissait tomber l'intuition individualiste, on pourrait imaginer que ce système opère de façon parfaitement distribuée, ce qui permettrait une telle

extension.

3.5 Haugeland

Le critère d'Haugeland (1993) pour déterminer l'extension de l'esprit est en fait originalement celui qu'Herbert Simon (1996) employait pour déterminer l'extension des composantes d'un système. La question qui intéresse Simon est de savoir comment on décompose un système en composantes plus petites pour faciliter l'analyse. Sa réponse se base sur cette observation :

In hierarchic systems, we can distinguish between the interactions among subsystems, on the one hand, and the interactions within subsystems— i.e., among the parts of those subsystems—on the other. The interactions at the different levels may be, and often will be, of different orders of magnitude. (Simon, 1996, 209)

L'idée est donc de découper les frontières de composantes à l'intérieur desquelles il y a beaucoup d'interaction, et en dehors desquelles il y en a peu – c'est une question d'intelligibilité. Haugeland propose d'appliquer ce principe au système cognitif. Pour ce faire, il distingue donc entre des axes d'interactions à « haut débit » (*high bandwidth*) par lesquelles s'échangent de grandes quantités d'information et les axes à « bas débit », où il en passe beaucoup moins. Or Haugeland argue que l'organisme et l'environnement sont dans une relation qu'il appelle d'« intimité » en ceci que l'interaction entre l'un et l'autre est tellement riche qu'il devient difficile de les distinguer. Aussi, le débit d'interaction ne se réduit pas lorsqu'on sort du cerveau et qu'on entre dans les interactions de l'organisme avec l'environnement – Haugeland (1993, 228) croit même qu'il augmente peut-être. Chose certaine, on reste dans le haut débit. D'où ceci, selon Haugeland, que l'esprit doit s'étendre.

À l'instar du critère de Clark, le critère d'Haugeland produit des frontières dynamiques (en ceci qu'elles changent selon ce que fait l'organisme), mais déterminées. La nature de l'heuristique de Simon demande de telles frontières, sans quoi il est

difficile de déterminer si les interactions sont à l'intérieur de la composante ou si elles se font avec l'extérieur. Par ailleurs, Haugeland s'accorde avec Simon pour dire qu'une interaction relativement intense tend à signifier une relative proximité spatiale dans les systèmes biologiques et physiques. Le système cognitif des organismes s'étend donc de façon approximable par un solide, et on devrait donc retrouver nos frontières.

Il y a cependant une ambiguïté chez Haugeland qui apparaît lorsqu'il parle de l'intelligence des organisations. Les individus qui les composent sont eux-mêmes en interactions à haut débit, de même qu'avec l'environnement et les objets qu'ils partagent (Haugeland, 1993, 236). Dans ce contexte, le sens des concepts ou des actions n'a pas à émerger dans la tête de quelqu'un – au contraire, pour Haugeland, l'intelligence vit dans un monde de sens, et non pas l'inverse. Aussi, comme une institution n'est pas, dans le sens qu'emploie l'auteur, un système biologique ou physique, elle n'a pas non plus à posséder la proximité spatiale des systèmes biologiques (Haugeland, 1993, 214). On n'a donc pas non plus, dans ce cas, de frontières au sens qui nous intéresse.

Dans le cas d'un critère comme celui de Haugeland, l'intuition individualiste n'est pas absolument nécessaire à la présence de frontières, puisqu'il suffit que le système identifié soit suffisamment compact – ce qui semble pouvoir être le cas de la plupart des systèmes physiques ou biologiques. En revanche, s'il avait adopté l'intuition individualiste, on aurait eu un système biologique – et donc on aurait aussi eu des frontières.

3.6 Rowlands

Dans *The New Science of the Mind* (2010b), Rowlands introduit lui aussi son propre critère, lequel est censé représenter la pratique en sciences cognitives, et se présente comme condition suffisante, mais non nécessaire, pour qu'un processus soit cognitif :

A process P is a cognitive process if :

1. P involves *information processing*—the manipulation and transformation of information-bearing structures.
2. This information processing has the *proper function* of *making available* either to the subject or to subsequent processing operations information that was, prior to this processing, unavailable.
3. This information is made available by way of the production, in the subject of P , of a *representational* state.
4. P is a process that *belongs* to the *subject* of that *representational state*. (Rowlands, 2010b, 110)

La quatrième condition, sur laquelle Rowlands s'attarde sur trois chapitres, est probablement la plus importante pour son projet. Or, son point focal étant l'appartenance au sujet, cela laisse supposer l'influence de l'intuition individualiste. Cependant, pour faire ce jugement, encore faut-il comprendre ce que Rowlands entend par « sujet » – il s'avère que ce concept n'est pas sans ambiguïté.

D'une part, le « sujet » de Rowlands correspond à ce qu'on entend généralement par « personne » sous plusieurs aspects. Premièrement, lorsque le sujet est sujet d'états représentationnels, ce sont des états représentationnels de niveau *personnel*. Les processus sous-personnels, en effet, n'informent pas le sujet directement : ils le font en informant les états de niveau personnel, qui à leur tour se présentent au sujet (Rowlands, 2010b, 151). Deuxièmement, cette activité de révélation d'information au sujet, et qui est l'essence de la cognition pour Rowlands, est une activité *consciente* (Rowlands, 2010b, 164). Le sujet est donc le sujet de l'expérience phénoménale. Troisièmement, Rowlands décrit explicitement le sujet comme étant une « personne », et cette personne, comme un « *organisme* capable de détecter des changements dans son environnement et de modifier son comportement en fonction de ceux-ci » (Rowlands, 2010b, 146, emphase ajoutée). Un tel organisme, qui vit dans son environnement, se distingue aussi de celui-ci, et forme donc un individu.

D'autre part, cependant, Rowlands ne pose pas de restrictions sur ce qui peut constituer une telle entité, de sorte qu'il aborde même la possibilité qu'il s'agisse d'un groupe d'organismes (Rowlands, 2010b, 135). Cette possibilité reste cependant plutôt abstraite, et Rowlands ne précise pas comment on peut marier l'idée

d'un groupe comme sujet cognitif avec le fait que ce dernier doive être une personne consciente.

Certes, Rowlands (2010a) est plutôt favorable à l'idée d'une extension de la conscience. Aussi n'est-il pas impossible que des groupes possèdent leur propre conscience collective et atteignent ainsi le statut de sujet cognitif. Cependant, il n'en reste pas moins que les conditions que Rowlands pose pour atteindre ce statut semblent plutôt contraignantes pour l'accession d'un groupe à ce statut, et en excluent beaucoup *a priori*. Par exemple, un groupe de travail ou un banc de poissons ne sauraient vraisemblablement pas répondre aux conditions nécessaires pour être organismes cognitifs, puisqu'il n'y a pas de conscience commune à laquelle peuvent être révélées des représentations. En revanche, le personnage de tante Hillary, la colonie de fourmis parlante de *Gödel, Escher, Bach : An eternal golden braid* (Hofstadter, 1999) est le prototype même de ce que serait un groupe qui aurait atteint le statut de cognitif, puisqu'elle possède clairement une conscience qui traite des représentations de niveau personnel. Ainsi, les groupes qui sont acceptables comme sujets cognitifs chez Rowlands sont en fait les groupes qui ont toutes les caractéristiques d'une personne.

En ce sens, ce n'est pas parce que Rowlands considère la possibilité que des groupes soient des systèmes cognitifs qu'il n'exige pas que tout système cognitif soit un individu. De fait, pour qu'un groupe soit un tel système, il est impératif qu'il forme lui-même un individu conscient – une personne.

Aussi peut-on dire que le critère de Rowlands satisfait d'une certaine façon les conditions pour le deuxième indice. Certes, il n'est pas tout à fait exact de dire que Rowlands cherche au niveau sous-personnel le principe de l'unité de l'individu, puisqu'il propose plutôt une explication qui fait du niveau personnel le niveau par lequel doit passer l'information pour que les représentations qui la convoient s'intègrent à l'individu. Cependant, il n'en reste pas moins qu'il propose un principe à l'unité de l'individu. Et comme les représentations de niveau personnel sont tout de même réalisées au niveau sous-personnel (Rowlands, 2010b, 144), ce principe doit avoir une explication au niveau sous-personnel, même si Rowlands ne s'y

attarde pas.

La façon qu'a Rowlands d'expliquer l'unité de l'individu n'est pas sans laisser entrevoir une hiérarchie à l'intérieur du système cognitif. D'une part, il y a deux types de processus cognitifs sous-personnels : ceux qui rendent leurs représentations accessibles au niveau personnel directement, et ceux qui les rendent accessibles aux processus de la première catégorie. Du moment qu'il y a une partie non-négligeable des processus cognitifs sous-personnels qui appartiennent à la seconde catégorie, alors on a effectivement une hiérarchie à même le niveau sous-personnel. On peut donc dire qu'il y a un sens en lequel, sans que ce soit une nécessité logique, le critère de Rowlands induit une hiérarchie. D'autre part, même s'il n'y avait pas de hiérarchie au niveau sous-personnel, il n'en reste pas moins que la dépendance au niveau personnel remplirait les conditions pour qu'on puisse dire qu'il y a une hiérarchie qui indique l'influence de l'intuition individualiste : la dépendance au niveau personnel, pour Rowlands, indique une dépendance à l'unité qu'est l'organisme – et donc l'individu.

Concernant le statut moral des processus cognitifs étendus, Rowlands n'en glisse pas mot. Cependant, comme il se prononce sur l'extension de la conscience, on peut en extrapoler des conclusions concernant le statut moral des mécanismes qui la réalisent. En tout cas, ne pas accorder de statut moral particulier à quelque chose dont on sait qu'il possède une conscience demanderait une justification particulière. Aussi le critère de Rowlands semble-t-il remplir les conditions pour qu'on y retrouve les trois indices de l'intuition individualiste.

Qu'en est-il des frontières ? D'une part, il ne semble pas y avoir de milieu entre l'appartenance au sujet et la non-appartenance. Soit un processus transfère une représentation à un autre, soit il ne le fait pas. Par ailleurs, comme l'appartenance repose sur la possibilité de ce transfert, il faut que les connexions nécessaires soient présentes. Dès lors, non seulement peut-on s'attendre à des frontières (déterminées par l'ensemble des processus qui appartiennent au sujet à un instant donné), mais elles doivent circonscrire une région sans discontinuité. On a donc des frontières qui font une forme que l'on peut approximer à un solide.

Qu'advient-il si on tente d'enlever l'intuition individualiste? Le problème ici est que le critère de Rowlands repose assez explicitement sur le concept d'individu, puisque c'est en lui présentant des informations qu'un mécanisme s'avère cognitif. Si l'on enlève ce concept, on ne peut plus construire de critère du cognitif. Par ailleurs, à cause de la place qu'occupe l'individu dans le critère de Rowlands, un système qui ne possède pas de frontières discrètes et continues ne saurait être un système cognitif à son sens. L'absence de discrétude ou de continuité romprait en effet l'intégrité de l'individu qui est nécessaire pour former un sujet cognitif.

3.7 Conclusion

Sur les quatre critères étudiés, trois reposent sur l'intuition internaliste. Dans tous les cas, la présence de frontières dépend elle aussi de la présence de l'intuition individualiste, et cette dépendance est confirmée par l'absence de frontières telles qu'on les entend dans le seul critère qui ne l'implique pas.

Est-ce suffisant pour conclure que la présence de frontières est due à l'intuition individualiste? Certes, il peut sembler que l'échantillon que nous avons choisi n'est pas suffisamment grand. Cependant, le fait est qu'il y a peu de critères qui permettent le genre de tests auxquels nous nous sommes adonnés ici – comme on l'a vu au chapitre précédent, beaucoup de critères ne sont pas suffisamment théoriquement complets pour notre usage.

Par ailleurs, on pourrait penser que notre hypothèse n'est pas formulée avec suffisamment de détail – par exemple, l'intuition individualiste n'est peut-être pas formulée de façon suffisamment formelle, de sorte qu'il est parfois difficile de ne pas projeter ses attentes sur un contexte textuel par biais d'interprétation. Certes, la difficulté, ici, est grande : comme on l'a vu, l'intuition individualiste s'applique de diverses manières en divers contextes, et les concepts participants à son énonciation (« individu », « personnel », « appartenance », etc.) ont tendance à changer d'un contexte à l'autre. Cependant, il semble bien y avoir un schéma commun,

comme le montre la répétition des traits que nous avons mentionné – hiérarchie dans le cognitif, recherche de principe sous-jacent à l'unité de l'individu au niveau sous-personnel – et la présence de frontières.

De plus, en suivant la logique dans laquelle s'inscrivait l'intuition individualiste chez chacun des auteurs, on remarque que la présence de frontières est une conséquence directe des aspects de leurs conceptions du cognitif qui découlent de l'intuition individualiste. Cette observation est relativement prévisible : l'idée derrière notre hypothèse est que, dans le contexte des débats sur l'esprit étendu, la présence de frontières est la conséquence logique de l'intuition individualiste.

C'est certainement le cas des auteur·es que nous avons étudié·es ici. Généraliser à partir d'un petit échantillon peut comporter un certain risque, mais à notre connaissance, la littérature n'offre aucun exemple contradictoire. Nous considérerons donc notre hypothèse comme suffisamment appuyée pour procéder à la dernière partie de notre argumentation.

CHAPITRE IV

CONCLUSION

4.1 Bilan de l'argumentation

Dans les chapitres 2 et 3, nous avons tâché de démontrer que l'hypothèse H2 était vraie, et sommes arrivés à un résultat qui nous paraît satisfaisant. Cependant, bien qu'il puisse être intéressant en soi de savoir que, si les caractérisations de l'esprit dépeignent toutes un espace du cognitif délimité par des frontières, c'est parce qu'elles reposent sur l'intuition individualiste, le but véritable ici reste de savoir s'il est légitime que les frontières en viennent à être l'option par défaut.

Bien entendu, nous avons pris pour hypothèse (H3) que ce n'est pas le cas : l'intuition individualiste, croyons-nous, ne justifie pas la préférence pour une extension par frontières, parce qu'elle manque elle-même de justification. Plus précisément, nous croyons qu'aucune des propositions qui font l'intuition individualiste n'est justifiée : ni celle concernant l'isomorphie, ni celle concernant le caractère individuel de l'esprit. Pour appuyer notre hypothèse, nous présentons ici quelques arguments à l'encontre de l'une ou l'autre de ces deux propositions.

On peut distinguer parmi celles-ci les critiques faibles et les critiques fortes. Pour les premières, l'isomorphie n'est pas toujours une mauvaise stratégie. Certes, on ne peut affirmer qu'une caractéristique de niveau sous-personnel prend une certaine forme ou une certaine valeur sur la base d'une relation d'isomorphie avec une

caractéristique de niveau personnel, par exemple. Cependant, si on ne fait qu’une critique faible de l’isomorphie, on peut quand même employer cette dernière pour justifier la présomption par défaut qu’on peut avoir pour cette caractéristique, en l’absence de connaissances empiriques pertinentes. Une critique forte, en revanche, ne considérerait pas qu’elle soit en mesure de justifier une telle présomption.

Pour que l’intuition individualiste soit falsifiée, il suffit qu’une seule des propositions qui la compose soit mauvaise. Cependant, aux fins de notre argumentation, il suffit de montrer qu’il est *plausible* qu’elle soit fausse : en effet, une proposition incertaine ne saurait être une base suffisamment solide pour justifier une connaissance.

Par ailleurs, la critique faible et la critique forte lancent différentes prescriptions ici : la critique faible nous enjoint à vérifier les prétentions des suppositions émanant de l’intuition individualiste, alors que la critique forte vise à mettre le fardeau de la preuve sur le dos de celles et ceux qui adoptent ces suppositions. Les deux types de critiques permettent d’appuyer nos hypothèses, mais elles le font de façons différentes, qui prêtent des interprétations légèrement différentes à nos hypothèses. Ainsi, dire que l’intuition individualiste est arbitraire est vrai à la lumière d’une critique faible, mais ne signifie pas pour autant qu’il soit injustifié de l’employer par défaut. En ce qui nous concerne, nous croyons que l’on peut reposer sur une critique forte de l’intuition individualiste – cependant, il faut noter que même si le lecteur n’était pas convaincu par les arguments de cette critique forte, cela ne nous libérerait pas pour autant de la nécessité de vérifier les prétentions empiriques de l’intuition individualiste.

4.1.1 Isomorphie

Comme on l’a dit plus haut, plusieurs critiques ont été avancées pour dénoncer l’emploi de l’isomorphie pour inférer des informations sur une description sous-personnelle du cognitif à partir de descriptions du cognitif au niveau personnel.

Parmi celles-ci, nous en relèverons deux – une, plus empirique, inspirée des succès de programmes de recherche en sciences cognitives, et l’autre, plus spéculative, inspiré d’un argument de Wimsatt (2007).

La première critique concerne les observations faites sur les systèmes dynamiques. Comme nous l’avons noté au chapitre I, Hurley (1998) croit, à l’instar de Port et Van Gelder, Kelso, et autres, que l’isomorphie entre niveaux faillit d’abord pour des raisons empiriques : le travail fait avec les systèmes dynamiques nous montre que les effets émergents sont souvent très imprédictibles. L’affirmation précise ici est que ce cas se retrouve dans une grande partie, voire la majorité des systèmes dynamiques. Comme le cognitif serait, de par sa complexité, le système dynamique par excellence (e.g. Van Gelder, 1998), il serait plausible qu’il suive cette tendance.

Force est de noter ici que cela n’empêche pas que l’isomorphie puisse avoir un certain potentiel de prédiction, au sens où il y a peut-être plus d’isomorphie dans les systèmes dynamiques à niveaux de description multiples que ne le prédirait le hasard. En ce sens, il s’agit d’une critique faible. L’objectif ici est principalement de faire une mise au point concernant le fardeau de la preuve : comme les systèmes dynamiques ont tendance à rendre les prédictions entre niveaux indignes de confiance, il faut à chaque fois les accompagner d’une démonstration.

La seconde critique, que nous avons brièvement mentionnée en section 1.2.1, développe sur la première en tentant d’expliquer les observations qu’on y retrouve. Elle fait appel à un certain concept de complexité, tel que développé et défendu chez Wimsatt (2007, chp. 9). Lorsqu’on étudie un système, on peut le regarder de l’angle de différentes approches théoriques. À chacune de ces approches correspond un certain découpage de l’objet décrit qui produit l’ontologie dont se sert la description. Dans certains systèmes (Wimsatt donne l’exemple d’une roche de granite) les différentes approches produisent des découpages qui réutilisent plus ou moins les mêmes frontières, ajoutant ou enlevant des détails au besoin : ces approches, selon Wimsatt, sont descriptivement simples. Si les frontières du découpage ontologique ne coïncident pas entre approches théoriques, on dira que le système est descriptivement complexe (Wimsatt donne l’exemple d’une mouche,

qu'on décompose fort différemment si on veut décrire l'anatomie de ses organes, ses réactions biochimiques ou ses phases de développement).

À cette complexité descriptive, on doit ajouter une complexité interactionnelle : pour qu'un système soit simple interactionnellement, il faut que les approches théoriques découpent l'animal à ses articulations, pour employer la célèbre métaphore de Platon. Wimsatt souscrit au concept de « décomposabilité quasi-complète » de Simon (1996) : un système est quasi-décomposable en sous-systèmes lorsque ceux-ci se caractérisent par une quantité beaucoup plus grande d'interactions internes que d'interactions avec l'extérieur.

La supposition d'isomorphie ne vaut que lorsque deux approches théoriques, et donc deux descriptions, partagent le même objet : il y a des invariants dans celui-ci qui garantissent une certaine similarité de la forme dans la façon dont il est décrit dans l'une et l'autre perspectives. Cependant, cela demande que le découpage se fasse dans une certaine mesure aux articulations que prescrit l'objet. Or, si on se fie à l'exemple de la mouche, celles-ci ne contraignent pas toujours ce découpage dans un système biologique. De fait, le choix des caractéristiques décrites dépend essentiellement des interactions qu'une approche théorique veut représenter.

Les descriptions sont donc *a priori* indépendantes les unes des autres. Cependant, on peut penser qu'il y a une contrainte dans la formation des approches théoriques qui pourrait les encourager à se former d'une façon à produire des ressemblances dans la description. Par exemple, Wimsatt croit lire chez Simon des contraintes qui seraient favorables à la formation d'isomorphies : d'une part, les approches théoriques se formeraient de façon à traiter intégralement de sous-systèmes obtenus selon le critère de quantité d'interactions de Simon – autrement dit, elles se formeraient de façon à maximiser la simplicité interactionnelle – et, d'autre part, la décomposition en sous-systèmes chez Simon se ferait de façon récursive, formant une hiérarchie de sous-systèmes. Si les approches théoriques correspondent à un niveau dans cette hiérarchie, alors celle qui s'intéresse à un niveau se retrouve à hériter des découpages en sous-systèmes faits au niveau précédent. Il y a donc matière à isomorphie, ne serait-ce que dans l'ontologie.

Cependant, cette association entre approche théorique et découpage interactionnel ne fonctionne pas, selon Wimsatt. Certes, la nature a intérêt à produire des êtres modulaires, comme le montre la fable de *Tempus* et de *Hora* (Simon, 1996, 188-90), mais ces modules se forment comme parties agrégables, et non comme parties indépendantes. En conséquence, elles peuvent facilement être dépendantes des autres parties pour leur fonctionnalité, et remplir plusieurs fonctions en même temps. De plus, faisant un parallèle avec l'histoire du développement des composants microélectroniques, Wimsatt avance qu'il existe des contraintes d'optimisation qui encouragent les systèmes à distribuer la réalisation des fonctions, et à réaliser plusieurs fonctions dans un même sous-système. Par ailleurs, étant donnée cette interdépendance entre sous-systèmes, le nombre et la complexité des liens qui les lient les uns avec les autres devraient rendre suspectes les décompositions récursives en composantes de plus en plus spécialisées : si les parties sous-systèmes multiplient les connexions vers l'extérieur, il n'est pas sûr du tout que ce partitionnement hiérarchique constitue une bonne heuristique pour opérer la décomposition¹.

Il n'y a donc pas de raison pour que les systèmes biologiques, l'esprit en particulier, soient simples interactionnellement, et en conséquence, il n'y a pas non plus de raison particulière pour qu'il y ait isomorphie entre descriptions. S'il n'y a pas de contrainte dans le développement des systèmes biologiques et de nos outils pour les connaître qui favorise l'isomorphie, alors il est assez improbable que celle-ci se réalise. On a donc ici une critique forte.

4.1.2 Caractère individuel

L'autre caractéristique de l'intuition individualiste, c'est qu'elle considère l'esprit comme une substance synthétique, autonome et distincte du non-cognitif. Mal-

1. Dans un cas où les sous-systèmes sont hautement dépendants du reste du système, il est vrai qu'un algorithme de partitionnement classique (e.g. *k-means*) ne produira pas de lignes de démarcation congruentes si on divise en deux puis en quatre : un partitionnement hiérarchique risque donc de produire des regroupements peu optimaux.

gré l'ambiguïté de l'expression, appelons-la « caractère individuel ». Par « synthétique », on entend l'idée selon laquelle l'esprit possède un principe d'individuation, qui garantit son unité dans le temps et dans l'espace – bref, que les informations que traite l'esprit se rassemblent en un endroit. L'autonomie est l'intuition selon laquelle nos pensées nous appartiennent en propre, de même que les opérations de la pensée dans lesquelles elles sont impliquées. La qualité d'être distincte du non-cognitif fait en sorte qu'on considère l'esprit comme possédant son domaine propre, régi par des régularités et des lois qui le distinguent du reste du monde physique. Ces caractéristiques correspondent à celles que nous avons utilisées en section 3.2.4 pour produire les tests qui nous ont aidés à détecter l'intuition individualiste dans le corpus d'étude.

Ce caractère individuel est lié à la notion d'un individu prototypiquement humain. Celui-ci est le lieu par excellence de la cognition – l'étalon à partir duquel on peut juger toute cognition. C'est donc dire que si la cognition de groupe, par exemple, est possible, ce n'est que parce qu'elle part d'individus et se fait pour leur profit, de telle sorte que, pour produire une analyse idoine d'un système cognitif de groupe, il faudrait le réduire aux contributions et intérêts individuels. Dans ce cadre, le caractère individuel devrait donc être compris comme un phénomène de niveau personnel.

Une caractéristique du caractère individuel est qu'il est censé refléter l'unité synthétique du sujet telle qu'elle apparaîtrait dans l'expérience phénoménale – caractéristique qu'il hérite de l'intuition individualiste en général. Aussi le caractère individuel peut-il être attaqué de deux façons : on peut vouloir nier l'unité synthétique du sujet phénoménal, ou on peut vouloir nier que cette unité phénoménale représente adéquatement le cognitif au niveau personnel.

Expérience unifiée ?

On peut d'abord se demander si le sujet est vraiment « unifié » dans l'expérience. Dans *Action in Perception*, Noë (2004) part de la critique de Hurley (2002) sur le schéma entrée-sortie du système cognitif pour produire une conception positive de la perception, qui veut que celle-ci soit une forme d'action. De fait, que la perception implique aussi des sorties d'information, sous forme de mouvements du corps, implique qu'elle demande aussi une certaine connaissance sensorimotrice. Les mouvements qui font partie de notre activité perceptive ne sont donc pas dépourvus d'intention – même si celle-ci est parfois implicite.

Pour Noë, ce ne sont pas les données de l'expérience qui, se ramenant à la conscience, se synthétisent en elles ; c'est plutôt le cognitif qui, s'adaptant à l'activité perceptive en cours, commande les actions perceptuelles pertinentes, met l'information entrante en contexte pertinent et complète les données manquantes avec les modèles déjà acquis à travers l'expérience sensorimotrice préalable. Bref, la perception est ce qui émerge de l'interaction entre l'organisme et l'environnement. Dans une telle conception de l'expérience phénoménale, la continuité du sujet ne s'inscrit pas dans sa permanence, mais plutôt dans sa capacité à se modeler pour répondre aux besoins de situation, en allant chercher les données disponibles les plus pertinentes à l'activité en cours.

Cette idée du sujet modelable est aussi présente chez Fine (2010) :

Some psychologists refer to whatever self is in current use – the particular self-concept chosen from the multitudes – as the active self. As the name implies, this is no passive, sloblike entity that idles unchanging day after day, week after week. Rather, the active self is a dynamic chameleon, changing from moment to moment in response to its social environment. Of course, the mind can only make use of what is available – and for each of us certain portions of the self-concept come more easily to hand than do others. But in all of us, a rather large portion of the Wardrobe of Self is taken up with the stereotypical costumes of the many social identities each person has (New Yorker, father, Hispanic American, vet, squash player, man). Who you are at

a particular moment – which part of your self-concept is active – turns out to be very sensitive to context. While sometimes your active self will be personal and idiosyncratic, at other times the context will bring one of your social identities hurtling towards the active self for use. With a particular social identity in place, it would not be surprising if self-perception became more stereotypic as a result. In line with this idea, priming gender seems to have exactly this effect. (Fine, 2010, 24-5)

Ces stéréotypes nous aident à choisir un schéma de comportement pertinent au contexte d'activité que nous vivons. Ils nous amènent à adopter certains comportements, et à les enligner avec un certain modèle. Cependant, ils nous amènent aussi à nous concevoir nous-mêmes de façons différentes – à nous attribuer certaines caractéristiques dans certains contextes, et leurs opposées dans d'autres.

Or ces stéréotypes qui modèlent nos actions, notre conception de nous-mêmes et sans doute le contenu de notre expérience sont des concepts que l'on acquiert dans notre environnement, à travers les représentations qui sont convoyées par notre entourage ou par les médias, par exemple. Bref, pour Fine, le soi, tel que l'individu le vit et le performe, est non seulement construit (notamment de l'extérieur par l'environnement social), mais cette construction aboutit à plusieurs identités que l'on revêt successivement selon l'occasion.

Ces façons de concevoir l'être humain ne suggèrent pas une synthèse de l'expérience pour un individu, mais plusieurs synthèses qui se font selon des modes adaptés au contexte du moment. L'information accessible à un sujet dépend alors de ce dernier. Dans un tel paradigme, attribuer un désir à l'individu plutôt qu'à l'identité ou au sujet contextualisé dans sa situation d'action ou de perception relève d'une simplification qui ne rend peut-être pas compte de la richesse de la vie cognitive et phénoménale de l'individu.

De plus, la synthèse d'informations disponibles au sujet de l'expérience ne se fait pas au niveau d'un soi central qui ferait office d'une sorte de poste de commande, mais émerge plutôt des interactions entre l'organisme et le monde. En ce sens, le sujet n'est pas autonome : il se construit en bonne partie au contact de l'extérieur,

allant jusqu'à lui emprunter des modes de fonctionnement.

Bref, il n'est pas du tout évident que l'expérience ait la forme que lui confère l'intuition individualiste, même s'il est vrai que la tradition phénoménologique, et en particulier celle qui a influencé les sciences cognitives, est restée attachée à l'autonomie et l'unité du sujet cartésien.

Theiner et l'ouverture à l'esprit collectif

Ceci dit, si la structure de l'expérience devait adéquatement être représentée par le caractère individuel, cela ne signifierait pas pour autant que cette dernière représente adéquatement le cognitif – comme on l'a dit plus haut, il est très possible que notre esprit fonctionne d'une façon radicalement différente à ce que suggère l'expérience que nous avons de notre cognition.

Dans « Recognizing Group Cognition », Theiner *et al.* (2010) s'affairent à montrer que la cognition peut se produire au niveau d'un groupe d'individus, comme des êtres humains. Pour eux, la cognition de groupe, du moins celle qui les intéresse, n'est pas réductible à la somme des cognitions des individus. Cependant, il ne s'agit pas non plus d'une forme de cognition ayant tous les attributs d'une cognition individuelle, mais se réalisant sur un groupe (comme c'était le cas chez Rowlands) : au contraire, pour Theiner et ses collaborateurs, la cognition de groupe a toutes les chances d'opérer de façon bien différente de celle qu'on retrouve chez les individus, et d'avoir des biais et des efficacités différentes pour différents types de problèmes.

Beaucoup de philosophes, notamment parmi celles et ceux qui s'intéressent à la question de l'extension de l'esprit, ont de fortes intuitions à l'encontre de la possibilité de la cognition de groupe – cela inclut, bien entendu, les détracteurs de l'esprit étendu, (e.g. Adams & Aizawa), mais aussi par plusieurs de ses défenseur·es (e.g. Clark). L'hypothèse de Theiner *et al.* est que l'attente des défenseur·es d'une

cognition exclusivement individuelle repose en partie sur une confusion quant à l'attribution du fardeau de la preuve. Ceux-ci considèrent en effet que l'existence de la cognition individuelle est évidente – par exemple, chacun de nous vit l'expérience de sa propre pensée – cependant, on ne posséderait pas de telle évidence pour un autre type de cognition. Dès lors, quiconque prétend en proposer une devrait produire une évidence de l'existence de celui-ci, faute de quoi il le risque de se produire un « appendice nomologique » (*nomological dangler*).

Pour renverser le fardeau de la preuve, Theiner *et al.* font appel au concept d'émergence tel que défini par Wimsatt (1985), c'est-à-dire l'échec d'agrégativité d'un niveau à l'autre. L'émergence est souvent appelée pour décrire une situation dans un système complexe où le niveau supérieur possède des qualités qui ne lui sont propres – qui ne se retrouvent ni ne s'expliquent dans aucun des constituants du système au niveau inférieur. La cognition de groupe qui intéresse les auteurs est un phénomène émergent en ce sens.

Un phénomène parallèle, pour Theiner *et al.*, est le phénomène de « démergence » : une situation où les structures de haut niveau non seulement ne possèdent pas certaines caractéristiques des structures de niveau inférieur, mais encore sont constitutionnellement incapables de les posséder. Pour les partisans d'une cognition strictement individuelle, la cognition serait donc un phénomène démergent : elle est possible à un niveau, mais le niveau supérieur ne peut la réaliser. Un tel scénario est pourtant improbable pour Theiner *et al.* : si les ressources à un niveau sont disponibles pour produire un phénomène émergent donné, comment ne le seraient-elles pas à un niveau d'organisation supérieur ? Aussi, si on doit faire une supposition, on doit conjecturer que si les ressources pour produire la cognition sont disponibles au niveau individuel, elles doivent l'être aussi au niveau du groupe. Le fardeau de la preuve tombe donc sur les épaules des défenseurs d'une cognition uniquement individuelle.

Ayant montré que la cognition de groupe est possible, Theiner et ses collaborateurs enchaînent sur quatre cas de cognition de groupe qui font office d'exemples, et montrent ainsi qu'elle est non seulement possible, mais qu'elle s'observe. Le

premier est celui de l'« *Active Walker* » : un modèle computationnel représentant comment les individus, en répondant à leurs propres problèmes de transport d'un point à un autre, produisent une solution à un problème communautaire, à savoir celui de produire un réseau de chemin efficace. Le second montre comment, dans un problème simple de coloration des nœuds d'un réseau, des agents en groupe performant mieux lorsque l'harmonisation des décisions est laissé aux dynamiques de groupe plutôt qu'à l'intelligence des agents individuels. Le troisième montre comment la division du travail s'acquiert, même sans communication entre les agents, dans la résolution d'un problème très simple. Enfin, le quatrième porte sur la remémoration collective – dont les performances seraient supérieures à la somme des performances des individus lorsque ces derniers savent comment interagir de façon à former ensemble un *système mémoriel transactif* (*transactive memory system* – cf. Wegner *et al.*, 1985).

Ces différents cas de figure exemplifient l'adaptation mutuelle des individus d'un groupe, qui se traduit au niveau du groupe par des conséquences qui peuvent être interprétées comme des solutions sous contraintes de sélection. Autrement dit, l'influence mutuelle des individus rend impossible de traiter quiconque comme étant le « cogniseur » ou la « cogniseuse ». Aussi peut-on avancer que la cognition de groupe est effective dans ces exemples.

Mais surtout, ces cas sont choisis pour montrer qu'il existe des cognitions (même humaines) qui ne sont pas représentables par un paradigme de cognition individuelle. On voit par ailleurs que de telles occurrences se produisent constamment : d'une part, comme le soulignent Theiner *et al.*, l'homme est un animal particulièrement grégaire, et d'autre part, les exemples choisis pourraient modéliser des comportements quotidiens. La cognition individuelle semble donc mal outillée pour représenter la cognition humaine.

4.1.3 Bilan des hypothèses

En somme, on peut dire qu'il est improbable que l'intuition individualiste ait la légitimité nécessaire pour justifier une extension du cognitif par frontières, car les deux piliers sur lesquels elle repose sont très vulnérables. D'une part, rien dans la façon dont on produit différentes approches théoriques pour expliquer un système dynamique ne suggère que celles-ci doivent produire des descriptions isomorphes, et l'expérience empirique semble confirmer cette prédiction ; l'inférence par isomorphie qu'emploie l'intuition individualiste semble donc compromise. D'autre part, le caractère individuel de l'esprit, qui devait nous informer sur la forme que l'on doit induire à la cognition au niveau sous-personnel, semble lui aussi suspect. Premièrement, certes, on pourrait penser qu'il provient de l'expérience phénoménale que le sujet a de lui-même, mais celle-ci pourrait tout aussi bien être perçue autrement. Deuxièmement, même si la cognition individuelle prenait cette forme, cela ne voudrait pas dire que toutes les cognitions y ressemblent – et on a vu dans l'exemple de la cognition de groupe un cas où la cognition prend une forme fort différente.

A-t-on démontré que l'intuition individualiste était fautive ? À cela, il faut répondre par la négative. L'intuition individualiste peut nous apparaître ici comme un pigeon d'argile, mais les failles que nous venons de souligner peuvent peut-être être colmatées. Certes, il ne semble pas y avoir isomorphie entre niveaux pour la plupart des systèmes dynamiques, mais peut-être y a-t-il quelque chose de particulier dans le fonctionnement de la cognition qui permettrait à cette isomorphie d'émerger. De même, malgré les doutes que soulèvent les approches énaclivistes comme celle de Noë et les approches « garde-robe » du soi comme celle de Fine, il est encore possible d'interpréter la structure du soi phénoménal comme revêtant un caractère individuel. Enfin, le caractère individuel est peut-être représentatif de la cognition de niveau personnel : l'existence de la cognition de groupe comme un phénomène distinct de la cognition individuelle ne fait certainement pas l'unanimité, et les propositions comme celles de Theiner sont considérées comme non orthodoxes – bien que ce pourrait précisément être parce qu'elles vont à l'encontre de l'intuition individualiste.

Cependant, l'hypothèse H3 ne suggère pas que l'intuition individualiste soit fausse, mais qu'elle n'est pas en mesure de justifier que l'extension de l'esprit prenne nécessairement une forme qui soit ceignable par des frontières. L'idée n'est donc pas de démontrer que l'intuition individualiste ne correspond pas à la réalité qu'elle prétend représenter, mais de montrer qu'elle ne saurait, du moins dans l'état actuel, remplir les conditions nécessaires pour être à la hauteur des ambitions que lui prêtent les philosophes qui l'utilisent pour fonder leurs conceptions de l'esprit.

Pour comprendre quelles sont ces conditions, il faut comprendre le rôle que l'intuition joue dans la construction des concepts : elle nous donne les outils pour produire un concept préliminaire, légitime dans la mesure où notre système cognitif est capable de produire des concepts adéquats sans l'aide de la métacognition, mais pas pour autant incontestable. Par exemple, pour Haslanger (2006), l'intuition sert à produire ce qu'elle appelle le *concept manifeste*, c'est-à-dire le concept tel qu'on l'entend lorsqu'on l'utilise. À travers diverses expériences de pensée, on peut approfondir ce concept, et le clarifier. Cependant, ce concept, qui correspond à l'entendement d'une personne, n'est pas nécessairement celui qui correspond à sa pratique : il se peut qu'elle emploie le concept dans une situation où son sens devrait le rendre inapproprié. Dans ce cas, pour résoudre cette contradiction, trois options se présentent à elle : aligner sa pratique à son concept, employer un autre concept correspondant à sa pratique (Haslanger nomme un tel concept « *concept opératif* ») ou aligner son concept à sa pratique (le nouveau concept ainsi produit est appelé « *concept visé* »).

Dans le cas qui nous intéresse, le concept serait celui du cognitif, et le concept manifeste serait celui d'une cognition individuelle, informée par l'intuition individualiste. Cependant, comme notre usage de ce concept est appliqué dans des cas comme la cognition de groupe, la réponse des bactéries à leur environnement ou l'adaptation que produit la sélection naturelle, on est à chaque fois confronté à un ou plusieurs concepts opératifs qui divergent du concept manifeste. La résolution implique toujours une modification, soit de notre usage du concept, soit de notre entendement de celui-ci : il n'y a donc pas de résolution conservatrice qui pourrait valoir par défaut.

Autrement dit, l'intuition ne saurait être en soi une raison pour préférer le concept manifeste au concept opératif. La raison ne pourrait provenir que d'une justification de l'intuition elle-même. Or l'examen que nous avons fait dans les sections précédentes ne nous a pas permis de trouver une telle raison – mais bien plutôt des raisons de douter de sa véracité. S'il n'y a pas de raison connue qui justifie l'intuition individualiste, alors il n'y a pas non plus de raison connue qui lui confère la légitimité pour justifier un type de caractérisation de l'esprit. Dans la mesure où l'on admet que l'espace des raisons qui nous intéresse est celui, discursif, des raisons qui sont actualisées, on peut donc penser que l'hypothèse H3 est démontrée – avec ce bémol que, si on en venait à découvrir des raisons justifiant l'intuition individualiste, cet espace viendrait à changer, et l'hypothèse pourrait s'avérer fausse.

Sous cette réserve, en conjuguant les hypothèses H2 et H3, on peut donc en conclure qu'il n'y a pas de raison particulière de penser que l'esprit doit se réaliser sous une forme qui soit descriptible par des frontières, et donc que notre hypothèse principale, H1, est elle aussi justifiée.

4.2 Horizons

4.2.1 Sciences humaines et intuition individualiste

La réserve que nous avons envers l'hypothèse H3 prend une couleur particulière quand on la conjugue à la remarque que nous avons faite en section 1.1 concernant les sciences humaines. Les raisons que nous avons étudiées pour justifier ou infirmer l'intuition individualiste proviennent toutes des sciences cognitives ; or il se peut que l'on puisse trouver des raisons en étudiant la portée de l'intuition individualiste dans les sciences humaines. De plus, comme le remarque Haslanger (2006), ce ne sont pas seulement des considérations d'épistémologie ou de progrès scientifique qui doivent peser dans le choix de la définition ou de la caractérisation d'un concept, mais aussi parfois des considérations de progrès social.

Dans la mesure où la littérature que nous avons étudiée n'aborde pas, ou alors très peu et de façon très elliptique, les relations entre le concept du cognitif ou de l'esprit au niveau des sciences cognitives ou de la philosophie de l'esprit et son impact dans les autres sciences humaines, leur étude ne cadrerait pas directement dans la problématique que nous nous sommes donnée, puisque celle-ci se limite à une certaine littérature. Cependant, dans un autre contexte, l'exercice ne saurait être que bénéfique. D'une part, cela élargirait la pertinence des conclusions produites par ce mémoire et mises à jour par cette nouvelle étude, et d'autre part, cela pourrait permettre de trouver dans les sciences humaines des éléments pertinents à la caractérisation de l'esprit.

Si nous ne nous sommes pas engagés dans cette étude dans le présent mémoire, il n'en reste pas moins que l'on peut avoir des intuitions quant à ce que l'on risque de trouver en approfondissant du côté des sciences humaines. En particulier, nous croyons que si les autres disciplines des sciences humaines subissent elles aussi l'influence d'intuitions parentes ou similaires à l'intuition individualiste, les conclusions que nous avons produites ici ne seront pas démenties pour autant. En effet, il nous semble que les sciences humaines ont déjà engagé, à leur façon, leur critique de l'intuition individualiste, ou alors ont trouvé des manières de contourner ses limitations.

On peut trouver un exemple de cela dans la distinction entre responsabilité criminelle et responsabilité civile. La responsabilité criminelle porte sur le non-respect d'une loi : en ce sens, c'est l'institution publique (la Couronne, l'État, etc.) qui fait la prosécution, et le jugement porte sur l'acte lui-même. En contrepartie, la responsabilité civile porte sur le non-respect de normes et de conventions morales entre deux partis dans le contexte social dans lequel ils évoluent (Duff, 2013). Dès lors, lorsque l'on juge la capacité d'une personne à remplir ses obligations, légales ou morales, envers la société ou envers autrui, les capacités cognitives de l'individu sont évaluées de façons très différentes. Pour la responsabilité criminelle, cette capacité (ou cette incapacité) s'évalue en termes de la santé mentale du défendeur ou de la défenderesse – c'est pourquoi on dira que la responsabilité criminelle pour faute est défectible par excuse (Coleman et Mendlow, 2010). En

revanche, en responsabilité civile, une faute n'est pas défectible par excuse : on peut donc être fautif en dépit d'une santé mentale gravement altérée au moment des événements, ou même en l'absence de tout manquement moral.

On considère ici que la responsabilité, en droit comme ailleurs, est intimement liée à la notion d'esprit ou de système cognitif : en effet, seuls des systèmes cognitifs peuvent être tenus responsables, parce que la capacité de comprendre un acte – et en particulier d'évaluer ses conséquences – n'est donnée qu'à un tel système. Inversement, le degré de responsabilité varie énormément selon les capacités cognitives de l'organisme, mais, généralement, tout système cognitif un peu élaboré a le potentiel d'être responsable de quelque chose dans une certaine mesure, du moment qu'il détermine lui-même la finalité de son comportement. La responsabilité porte donc sur un système cognitif : si notre conception du cognitif est strictement individuelle, par exemple, elle ne pourra porter que sur l'individu.

Prenons par exemple l'arrêt *Miller v Jackson* (1977) : dans cette décision, Lord Denning et ses collègues évaluent un cas de responsabilité civile concernant les dommages causés par des balles de cricket qui se perdent occasionnellement dans la cour de résidences avoisinant un terrain de cricket. Celui-ci constitue un appel à un autre jugement, dans lequel le juge reconnaît le club de cricket responsable des dommages, et lui ordonne de cesser ses activités dans le terrain de cricket en question. Cependant, le terrain est beaucoup plus ancien que les résidences qui le bordent. En ce sens, sont aussi responsables du dommage les promoteur·ses qui ont construit les habitations, les autorités municipales qui ont permis ces constructions et les acheteur·ses qui n'ont pas pris compte des conséquences prévisibles de la présence du terrain de cricket à côté de leur propriété.

La responsabilité civile se voit ainsi souvent en face d'un problème de responsabilité partagée. Pourtant la responsabilité doit être assumée quelque part. Dans certains cas, la santé de population en dépend (e.g. *Donoghue v Stevenson* 1932). La solution consiste à rendre les individus responsables de comportements qui portent en fait sur leur entourage social. Par exemple, il n'est pas clair que les habitant·es des propriétés bordant le terrain de cricket aient été informé·es du fait

que ce terrain servait au cricket, ni du fait que lorsqu'on utilise un terrain de cette grandeur, il est inévitable que des balles de cricket soient projetées à l'extérieur de temps en temps. Cette responsabilité revient aux propriétaires précédents, ou aux promoteur·ses qui leur ont vendu leur habitation.

Dans le cas du droit criminel, la loi prescrit généralement assez clairement quelle entité peut lui contrevenir. En revanche, dans le cas de la responsabilité civile, la norme étant indéterminée et le jugement ayant pour but de réparer un tort, les contraintes sur l'entité responsable sont plus souples. On se trouve donc souvent devant des responsabilités collectives ou partagées, qui ne cadrent pas avec la conception individuelle de la responsabilité qui se trouve dans les lois. Cependant, comme les options de compensation des juges sont limitées, ils sont obligés de faire porter la responsabilité collective sur un ou plusieurs individus impliqués. En ce sens, une des caractéristiques les plus contre-intuitives du droit civil, à savoir le fait que l'on puisse être responsable sans avoir commis la moindre faute individuelle, semble être une adaptation à la contradiction entre une conception individuelle de la responsabilité (et donc aussi de la cognition) et une réalité de responsabilités partagées. Autrement dit, la responsabilité civile est un concept de responsabilité non-individuelle (quoique centrée sur l'individu) qui repose sur une conception non-individuelle du cognitif alors que la responsabilité criminelle est individuelle et repose sur une conception correspondante du cognitif.

4.2.2 Contribution aux sciences cognitives

Ceci dit, quelle est l'importance de notre travail pour les sciences cognitives, en dehors du domaine d'intérêt tel que défini par le corpus de référence ?

D'entrée de jeu, il faut reconnaître qu'en elle-même, notre hypothèse principale, H1, a peu d'impact direct dans les sciences cognitives en dehors du corpus de référence. En effet, si on se limite à cette seule hypothèse, force est de reconnaître qu'elle porte sur l'extension de l'esprit : si une théorie ou un programme

de recherche est agnostique quant à cette question – et la majorité des sciences cognitives le sont – alors, *a priori*, H1 a peu d'importance pour celle-ci ou celui-ci. La même chose peut d'ailleurs être dite de l'hypothèse H2 : l'association entre frontières et intuition individualiste suppose la présence de frontières, et donc un intérêt pour la question de l'extension.

Ceci dit, H1 et H2 contribuent surtout aux sciences cognitives dans la mesure où elles font avancer le débat sur la caractérisation et l'extension de l'esprit. De fait, en ouvrant les portes à de nouvelles manières de déterminer l'extension du cognitif, elles peuvent contribuer à l'avènement de caractérisations de l'esprit qui sont assez peu orthodoxes, et qui pourraient ainsi avoir un impact important dans les disciplines pertinentes.

L'hypothèse H3, en revanche, a beaucoup plus de portée : bien que l'on n'ait pas tenté d'évaluer son impact dans les sciences cognitives, il est probable qu'elle s'étende à de nombreuses questions. En effet, l'intuition individualiste est pertinente pour toutes les questions qui peuvent porter ou non sur la cognition individuelle, et pour lesquelles cette distinction est pertinente. Cela exclut certes les travaux plus spécifiques – portant, par exemple, sur la mémoire individuelle humaine, ou sur le potentiel d'apprentissage d'un réseau de neurones – mais sinon, toutes les disciplines sont susceptibles d'être influencées par l'intuition individualiste. De fait, comme le montre notre exemple avec la responsabilité juridique, il est probable que l'hypothèse H3 soit pertinente bien au-delà des sciences cognitives, dans l'ensemble des sciences humaines.

En ce sens, la contribution de ce mémoire aux sciences cognitives en général n'est pas seulement dans la conclusion qu'il produit, mais aussi dans les concepts qu'il développe pour y parvenir. On a défini l'intuition individualiste pour un usage particulier qui s'applique à un corpus très réduit ; cependant – et ce même s'il semble que l'intuition individualiste, dans un autre corpus, devrait produire d'autres effets, et qu'en conséquence, on devrait adapter notre définition en changeant de corpus – il s'agit d'une base conceptuelle qui peut servir à le détecter dans d'autres corpus. Comme elle reste (sous toutes réserves) sans justification, il serait gran-

dement utile d'examiner ses influences ailleurs dans les sciences cognitives et en général dans les sciences humaines afin de corriger les erreurs qu'elle pourrait y induire.

Ceci dit, on doit dire « sous toutes réserves », parce que l'hypothèse H3 n'a pas fait ici l'objet d'une étude systématique. D'une part, le fait que l'intuition individualiste n'ait pas encore fait l'objet d'une défense et d'un débat rend tout projet de révision massive prématurée. Dans la même veine, l'étude de l'hypothèse H3 dans ce mémoire n'a rien de systématique : il est très possible que des arguments puissants aient passé sous le radar – d'autant plus que l'intuition individualiste étant implicite, il est difficile de retracer ces arguments. D'autre part, le fait que nous n'ayons pu étudier les justifications pour celle-ci dans les sciences humaines fait en sorte que notre traitement de H3 est insuffisant pour en généraliser la portée.

Tout en arrivant avec une conclusion qui nous semble solide à l'intérieur des paramètres donnés, ce mémoire se conclut donc sur cette double ouverture : d'une part, vers d'autres solutions à la question de l'extension, et d'autre part, vers un renforcement de l'hypothèse H3 et une étude de l'intuition individualiste dans l'ensemble des sciences cognitives et des sciences humaines.

RÉFÉRENCES

Donoghue v Stevenson, 1932 UKHL 100.

Miller v Jackson, 1977 QB 966.

Adams, F. et Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14(1) : 43–64.

Adams, F. et Aizawa, K. (2008). *The Bounds of Cognition*. Malden, MA : Wiley-Blackwell.

Adams, F. R. et Aizawa, K. (2005). Defending Non-Derived Content. *Philosophical Psychology*, 18(6) : 661–669.

Aizawa, K. (2010, 16 décembre). Menary on Lack of Consensus. [Billet de blogue]. Récupéré de *The Bounds of Cognition* <http://theboundsofcognition.blogspot.com/2010/12/menary-on-lack-of-consensus.html>.

Allen-Hermanson, S. (2012). Superdupsizing the Mind : Extended Cognition and the Persistence of Cognitive Bloat. *Philosophical Studies*, 158(1) : 1–16.

Anderson, J. (2008). Neuro-Prosthetics, the Extended Mind, and Respect for Persons with Disability. *The Contingent Nature of Life* 259–274.

Anscombe, G. E. M. (1957). *Intention*. Cambridge, MA : Harvard University Press.

Aristote (2007). *Rhétorique*. (Pierre Chiron, trad.). Paris : Flammarion.

- Baggini, J. (2008). A Piece of iMe : An Interview with David Chalmers. *The Philosophers' Magazine*, 43.
- Barker, M. J. (2010). From cognition's location to the epistemology of its nature. *Cognitive Systems Research*, 11(4) : 357–366.
- Beer, R. D. (1990). *Intelligence as adaptive behavior : an experiment in computational neuroethology*. San Diego, CA, USA : Academic Press Professional, Inc.
- Blake, A. et Yuille, A. A. L. (1992). *Active Vision*. Artificial Intelligence Series. Cambridge, MA : MIT Press.
- Block, N. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, 9 : 261–325.
- Block, N. (1990). Inverted earth. *Philosophical perspectives*, 4 : 53–79.
- Braddon-Mitchell, D. et Jackson, F. (2006). *Philosophy of Mind and Cognition : An Introduction*. Malden, MA : Wiley.
- Bradley, F. H. (1895). In What Sense Are Psychological States Extended? *Mind*, 4(14) : 225–235.
- Brentano, F. (1973). *Psychology from an empirical standpoint*. (Antos C. Rancurello, D.B. Terrel, et Linda McAlister, trad.). London : Routledge & Kegan Paul.
- Brooks, R. A. (1986). A robust layered control system for a mobile robot. *Robotics and Automation, IEEE Journal of*, 2(1) : 14–23.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, 47(1-3) : 139–159.
- Buller, T. (à paraître). Neurotechnology, Invasiveness and the Extended Mind. *Neuroethics*.

- Cangelosi, A. et Harnad, É. (2001). The adaptive advantage of symbolic theft over sensorimotor toil : Grounding language in perceptual categories. *Evolution of Communication*, 4(1) : 117–142.
- Carnap, R. (1950). *Logical Foundations of Probability*. Chicago : University of Chicago Press.
- Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3) : 200–219.
- Chalmers, D. J. (2008). Foreword to Andy Clark’s *Supersizing the mind*. Dans *Supersizing the mind*. Oxford : Oxford University Press.
- Chisholm, R. M. (1957). *Perceiving : a philosophical study*. Contemporary philosophy. Ithaca, NY : Cornell University Press.
- Clark, A. (1998). *Being there : putting brain, body, and world together again*. Bradford Books. Cambridge, MA : MIT Press.
- Clark, A. (2006). Material Symbols. *Philosophical Psychology*, 19(3) : 291–307.
- Clark, A. (2008). *Supersizing the mind*. Oxford : Oxford University Press.
- Clark, A. (2009). Spreading the Joy? : Why the Machinery of Consciousness is (Probably) Still in the Head. *Mind*, 118(472) : 963–993.
- Clark, A. et Chalmers, D. (1998). The extended mind. *Analysis*, 58(1) : 7.
- Clark, A. et Toribio, J. (1994). Doing Without Representing. *Synthese*, 101(3) : 401–31.
- Clowes, R. et Morse, A. F. (2005). Scaffolding Cognition with Words. Dans Berthouze, L., Kaplan, F., Kozima, H., Yano, H., Konczak, J., Metta, G., Nadel, J., Sandini, G., Stojanov, G., et Balkenius, C. (dir.), *Fifth International Workshop on Epigenetic Robotics*, volume 123 : 101–105. Lund : Lund University Cognitive Studies.

- Coleman, J. et Mendlow, G. (2010). Theories of Tort Law. Dans Zalta, E. N. (dir.), *The Stanford Encyclopedia of Philosophy*. (Édition automne 2010). Récupéré de <http://plato.stanford.edu/archives/fall2010/entries/tort-theories/>.
- Crane, T. (1998). Intentionality as the mark of the mental. *Royal Institute of Philosophy Supplements*, 43 : 229–251.
- Craver, C. (2007a). Constitutive Explanatory Relevance. *Journal of Philosophical Research*, 32 : 3–20.
- Craver, C. F. (2007b). *Explaining the Brain : Mechanisms and the Mosaic Unity of Neuroscience*. Oxford : Oxford University Press, Clarendon Press.
- Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA : MIT press.
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23) : 685–700.
- Davidson, D. (1970). Mental Events. Dans *Experience and Theory* (p. 79–101). Humanities Press.
- Degré, K. (2011). L'interprétation mécaniste des communautés et des écosystèmes. (Mémoire de maîtrise). Université de Montréal. Récupéré de *Papyrus*, dépôt institutionnel numérique de l'Université de Montréal <http://www.archipel.uqam.ca/3251/>.
- Dennett, D. C. (1989). *The Intentional Stance*. Cambridge, MA : MIT Press.
- Devitt, M. et Sterelny, K. (1999). *Language and Reality : An Introduction to the Philosophy of Language*. (2^e édition). Cambridge, MA : MIT Press.
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3(4) : 357–370.
- Dewey, J. (1916). *Essays in experimental logic*. Chicago : The University of Chicago Press.

- Drayson, Z. et Clark, A. (à paraître). Augmentation, Agency, and the Spreading of the Mental State. *Neuroethics*.
- Dretske, F. (1981). *Knowledge & the flow of information*. Cambridge, MA : MIT Press.
- Dretske, F. (1988). *Explaining behavior : Reasons in a world of causes*. Cambridge, MA : MIT Press.
- Duff, A. (2013). Theories of Criminal Law. Dans Zalta, E. N. (dir.), *The Stanford Encyclopedia of Philosophy*. (Édition été 2013). Récupéré de <http://plato.stanford.edu/archives/sum2013/entries/criminal-law/>.
- Dunbar, K. (1997). How scientists think : On-line creativity and conceptual change in science. Dans Ward, T. B., Smith, S. M., et Vaid, J. (dir.), *Creative thought : An investigation of conceptual structures and processes (p. 461–493)*. Washington, DC : American Psychological Association.
- Fayyad, U., Piatetsky-Shapiro, G., et Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3) : 37.
- Fine, C. (2010). *Delusions of gender : How our minds, society, and neurosexism create difference*. New York : WW Norton & Company.
- Fisk, J., Lackner, J. R., et DiZio, P. (1993). Gravitoinertial force level influences arm movement control. *Journal of neurophysiology*, 69(2) : 504–511.
- Fodor, J. (1989). *Psychosemantics*. Cambridge, MA : MIT Press.
- Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA : The MIT Press.
- Fodor, J. A. (2001). *The Mind Doesn't Work That Way : The Scope and Limits of Computational Psychology*. Bradford Books. Cambridge, MA : MIT Press.
- Gallagher, S. (2008). Are Minimal Representations Still Representations? *International Journal of Philosophical Studies*, 16(3) : 351–369.

- Gallagher, S. (2011). *The Socially Extended Mind*. Conférence donnée à la Freie Universität Berlin. Récupéré de http://www.hum.au.dk/semiotics/docs2/news_archive/2011/shaun-gallagher-masterclasses/the-socially-extended-mind.pdf.
- Gallagher, S. (2012). The Overextended Mind. *Versus*, 113 : 57–68.
- Gallagher, S. et Crisafi, A. (2009). Mental Institutions. *Topoi*, 28(1) : 45–51.
- Gallagher, S. et Miyahara, K. (2012). Neo-Pragmatism and Enactive Intentionality. Dans *New Directions in Philosophy and Cognitive Science : Adaptation and Cephalic Expression*. Basingstoke, UK : Palgrave Macmillan.
- Gigerenzer, G. et Brighton, H. (2009). Homo heuristicus : Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1) : 107–143.
- Goldin-Meadow, S. (2005). *Hearing gesture : How our hands help us think*. Cambridge, MA : Harvard University Press.
- Golonko, S. (2010, 14 octobre). Is it time to abandon the cognitive / non-cognitive distinction? [Billet de blogue]. Récupéré de *Notes from Two Scientific Psychologists* de <http://psychsciencenotes.blogspot.ca/2010/10/is-it-time-to-abandon-cognitive-non.html>.
- Harnad, É. (1990). The symbol grounding problem. *Physica D : Nonlinear Phenomena*, 42(1-3) : 335–346.
- Haslanger, S. (2006). What Good Are Our Intuitions? Dans *Proceedings of the Aristotelian Society Supplementary Volume*, volume 80 : 89–118. Malden, MA : Wiley Online Library.
- Haugeland, J. (1993). Mind Embodied and Embedded. Dans *Mind and Cognition : 1993 International Symposium (p. 233–267)*. Taipei, Taiwan : Academica Sinica.
- Heider, F. et Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2) : 243–259.
- Hofstadter, D. R. (1999). *Gödel, Escher, Bach : an eternal golden braid*. Art of Mentoring Series. New York : Basic Books.

- Hurley, S. (2010). Varieties of externalism. Dans Menary, R. (dir.), *The Extended Mind* (p. 101–154). Cambridge, MA : MIT Press.
- Hurley, S. L. (1998). Vehicles, Contents, Conceptual Structure and Externalism. *Analysis*, 58(1) : 1–6.
- Hurley, S. L. (2002). *Consciousness in action*. Cambridge, MA : Harvard University Press.
- Hutchins, E. (1991). The social organization of distributed cognition. Dans L.B. Resnick, J.M. Levine, S. T. (dir.), *Perspectives on socially shared cognition*. Washington, DC : American Psychological Association.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA : MIT press.
- Hutchins, E. (2011). Enculturating the Supersized Mind. *Philosophical Studies*, 152(3) : 437–446.
- Jacob, P. (2010). Intentionality. Dans Zalta, E. N. (dir.), *The Stanford Encyclopedia of Philosophy*. (Édition automne 2010). Récupéré de <http://plato.stanford.edu/archives/fall2010/entries/intentionality/>.
- Kahneman, D. et Frederick, S. (2002). Representativeness revisited : Attribute substitution in intuitive judgment. Dans Gilovich, T., Griffin, D., et Kahneman, D. (dir.), *Heuristics and biases : The psychology of intuitive judgment* (p. 49–81). Cambridge, UK : Cambridge University Press.
- Kaplan, D. M. (2012). How to Demarcate the Boundaries of Cognition. *Biology and Philosophy*, 27(4) : 545–570.
- Kirsh, D. (1995). The intelligent use of space. *Artificial intelligence*, 73(1) : 31–68.
- Kirsh, D. et Maglio, P. P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18(4) : 513–549.
- Lakatos, I. (1970). *The Methodology of Scientific Research Programmes : Philosophical Papers*, volume 1, chap. Falsification and the methodology of scientific research programmes (p. 8–101). Cambridge, UK : Cambridge University Press, Cambridge, UK.

- Liggins, J., Pihl, R. O., Benkelfat, C., et Leyton, M. (2012). The Dopamine Augmenter L-DOPA Does Not Affect Positive Mood in Healthy Human Volunteers. *PloS one*, 7(1).
- Machamer, P. K., Darden, L., et Craver, C. F. (2000). Thinking About Mechanisms. *Philosophy Of Science*, 67(1) : 1–25.
- McClelland, J. L., Rumelhart, D. E., et Hinton, G. E. (1986). Parallel distributed processing. Dans Rumelhart, D. E., McClelland, J. L., et PDP Research Group, C. O. R. P. O. R. A. T. E. (dir.), *The appeal of parallel distributed processing : explorations in the microstructure of cognition*, volume 1 (p. 3–44). Cambridge, MA : MIT Press. Récupéré de <http://dl.acm.org.proxy.bibliotheques.uqam.ca:2048/citation.cfm?id=104279.104284>.
- McGinn, C. (1995). Consciousness and space. *Journal of Consciousness Studies*, 2(3) : 220–230.
- McNeill, D. (2005). *Gesture and thought*. Chicago : University of Chicago Press.
- Mead, G. H. (1929). A pragmatic theory of truth. *University of California Publications*, 11(3).
- Mead, G. H. (1938). *The Philosophy of the Act*. Chicago : University of Chicago Press.
- Menary, R. (2007). *Cognitive integration : mind and cognition unbounded*. Basingstoke, UK : Palgrave Macmillan.
- Menary, R. (2009). Intentionality, Cognitive Integration and the Continuity Thesis. *Topoi*, 28(1) : 31–43.
- Menary, R. (2010). The holy grail of cognitivism : a response to Adams and Aizawa. *Phenomenology and the Cognitive Sciences*, 9(4) : 605–618.
- Meunier, J. G., Forest, D., et Biskri, I. (2005). Classification and categorization in computer assisted reading and analysis of texts. Dans Cohen, H. et Lefebvre, C. (dir.), *Handbook on categorization in cognitive science* (p. 955–978). Amsterdam : Elsevier.

- Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA : MIT Press.
- Millikan, R. G. (1996). Pushmi-Pullyu Representations. *Philosophical Perspectives*, 9 : 185–200.
- Millikan, R. G. (2000). Naturalizing Intentionality. Dans Elevitch, B. (dir.), *The Proceedings of the Twentieth World Congress of Philosophy*, volume 9 (p. 83–90). Charlottesville, VA : Philosophy Documentation Center.
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., et Handelsman, J. (2012). Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41) : 16474–16479.
- Noë, A. (2004). *Action in perception*. Cambridge, MA : MIT Press.
- Onishi, K. H. et Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719) : 255–258.
- Philippides, A., Husbands, P., Smith, T., et O’Shea, M. (2005). Flexible Couplings : Diffusing Neuromodulators and Adaptive Robotics. *Artif. Life*, 11(1-2) : 139–160.
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in cognitive sciences*, 11(1) : 37.
- Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19) : 699–711.
- Pylyshyn, Z. W. (1980). Computation and cognition : Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3(1) : 111–169.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA : MIT Press.
- Rorty, R. (1970a). Cartesian Epistemology and Changes in Ontology. Dans Smith, J. E. (dir.), *Contemporary American Philosophy*. New York : Humanities Press.
- Rorty, R. (1970b). Incommensurability as the Mark of the Mental. *Journal of Philosophy*, 67(12) : 399–424.

- Rowlands, M. (2006). *Body Language : Representation in Action*. Bradford Book. Cambridge, MA : MIT Press.
- Rowlands, M. (2009). Extended Cognition and the Mark of the Cognitive. *Philosophical Psychology*, 22(1) : 1–19.
- Rowlands, M. (2010a). Consciousness broadly construed. Dans Menary, R. (dir.), *The extended mind*. Cambridge, MA : MIT Press.
- Rowlands, M. (2010b). *The New Science of the Mind : From Extended Mind to Embodied Phenomenology*. Bradford Books. Cambridge, MA : MIT Press.
- Rupert, R. D. (2009). *Cognitive systems and the extended mind*. Oxford : Oxford University Press, USA.
- Schneider, S. (2011). *The Language of Thought : A New Philosophical Direction*. Cambridge, MA : MIT Press.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., et Simons, A. (1991). Ease of retrieval as information : Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61(2) : 195.
- Searle, J. R. (1958). Proper Names. *Mind*, 67(266) : 166–173.
- Searle, J. R. (1983). *Intentionality*. Cambridge, MA. : Cambridge University Press.
- Searle, J. R. (1984). Intentionality and its place in nature. *Synthese*, 61(1) : 3–16.
- Searle, J. R. (1992). *The rediscovery of the mind*. Representation and mind. Cambridge, MA : MIT Press.
- Simon, H. A. (1996). *The Sciences of the Artificial*. Cambridge, MA : MIT Press.
- Smart, J. J. C. (2011). The Mind/Brain Identity Theory. Dans Zalta, E. N. (dir.), *The Stanford Encyclopedia of Philosophy*. (Édition automne 2011). Récupéré de <http://plato.stanford.edu/archives/fall2011/entries/mind-identity/>.

- Sprevak, M. (2009). Extended Cognition and Functionalism. *Journal of Philosophy*, 106(9) : 503–527.
- Sprevak, M. (2010). Inference to the Hypothesis of Extended Cognition. *Studies in History and Philosophy of Science*, 41(4) : 353–362.
- Steels, L. (1999). *The Talking Heads Experiment*. Antwerpen : Laboratorium.
- Suchman, L. A. (1987). *Plans and situated actions : the problem of human-machine communication*. New York : Cambridge University Press.
- Sutton, J. (2004). Representation, Reduction, and Interdisciplinarity in the Sciences of Memory. Dans Hugh Clapin, Phillip Staines, P. S. (dir.), *Representation in Mind : New Approaches to Mental Representation*. Amsterdam : Elsevier.
- Thagard, P. (1995). Book Review : What Is Cognitive Science ? by Barbara Von Eckardt. *Philosophy of Science*, 62(2) : 345.
- Theiner, G., Allen, C., et Goldstone, R. L. (2010). Recognizing group cognition. *Cognitive Systems Research*, 11(4) : 378–395.
- Thelen, E. S. et Smith, L. B. (1996). *Dynamic Systems Approach to the Develop.* Cambridge, MA : MIT Press.
- Tversky, A. et Kahneman, D. (1973). Availability : A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2) : 207–232.
- Ullman, S. et Richards, W. (1984). *Image Understanding*. New York : Ablex Publishing Corporation.
- Van Gelder, T. (1990). Why Distributed Representation is Inherently Non-Symbolic. Dans *Konnektionismus in Artificial Intelligence Und Kognitionsforschung*. Berlin : Springer-Verlag.
- Van Gelder, T. (1995). What might cognition be, if not computation ? *The Journal of Philosophy*, 92(7) : 345–381.

- Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain sciences*, 21(05) : 615–628.
- Vogt, P. (2000). *Lexicon grounding on mobile robots*. Thèse de doctorat, Vrije Universiteit Brussel, Brussel.
- Von Eckardt, B. (1995). *What is cognitive science ?* Bradford Books. Cambridge, MA : MIT Press.
- Wegner, D. M., Giuliano, T., et Hertel, P. T. (1985). Cognitive interdependence in close relationships. Dans *Compatible and incompatible relationships* (p. 253–276). Berlin : Springer.
- Wheeler, M. (2005). Friends reunited ? : Evolutionary robotics and representational explanation. *Artificial Life*, 11(1-2) : 215–231.
- Wheeler, M. (à paraître). *Extended X : Recarving the Biological and Cognitive Joints of Nature*.
- Wimsatt, W. C. (1985). Forms of aggregativity. Dans *Human nature and natural knowledge* (p. 259–291). Berlin : Springer.
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings : piecewise approximations to reality*. Cambridge, MA : Harvard University Press.